

The promise of proteomics

Analysing the entire set of proteins of an organism is a far bigger challenge than anything in genomics. The technological obstacles and biological complexities require, for now, a steady approach to that necessary goal.

The inside of a cell is a crowded and dynamic place, where proteins are perpetually being created and discarded. Understanding the structures, interactions and functions of all of a cell's or organism's proteins is one of the grand goals of the post-genomic era, and has been given a disciplinary title of its own: proteomics. There are even some who want to develop a human proteome project. Is that premature, or even meaningful?

Proteomics, if defined as the study of many proteins simultaneously in order to understand the function of one restricted state of a cell, remains in its infancy. A pioneering example was published last month (*Nature* **402**, 147–154; 1999). That paper reported that, out of the 2,500 polypeptides in the cytosol of the bacterium *Escherichia coli*, only around 300 proteins use cylindrical (GroEL) molecules as chaperones to ensure that they are correctly folded. The analysis, using two-dimensional gels and database comparisons, positively identified more than 50 of these, and highlighted key structural features that determine these proteins' interactions with, or need for, 'chaperonins'.

As is described in this week's survey of the prospects for proteomics (see page 715), the now well-established two-dimensional-gel approach has many limitations, while the development of the more advanced technologies we can envisage — antibody and protein-array technologies that may deliver fast and parallel quantitative analyses of protein distributions — has a long way to go. Moreover, as the number of proteins identified in a cellular event grows, so too do the demands to validate those identifications and infer the proteins' activities in order to give meaning to the observations. New centres are springing up in many universities, bringing together the skills and technologies required to tackle such challenges. Alongside the embryonic technical state of proteomics, there is, critically, a shortage of people skilled in bioinformatics.

On top of these technical limitations is a more fundamental issue.

The number of genes in an individual human, as in any organism, is static and fixed. Given the much larger set of proteins produced by that organism at one time or another throughout its life, the goal of identifying the whole of the human proteome is a far bigger and more complex challenge. Indeed, there is no such thing as 'the' human proteome — it will differ significantly not only between individuals (much more than do their genomes), but also within one individual before and after, say, a millennium party.

So, should funding agencies be pouring money into some global strategy at this point? A boost now risks committing large sums to techniques that may soon be superseded. Moreover, there are already more than enough known molecular systems that need to be targeted in a focused way. Such arguments were initially raised against the Human Genome Project and with hindsight were a distraction. But here and now, with respect to proteomics, they are apt. Collaborations between many centres can add value by concentrating on the proteins involved in aspects of cell function and development selected for their wide relevance across the kingdoms of life or their particular relevance to human health. The fact that the pharmaceutical industry is excited about the dawning prospects of using proteomics to help identify new drug targets, but has yet to invest substantially, reflects the wisdom of a steady approach to the growth of proteome biology.

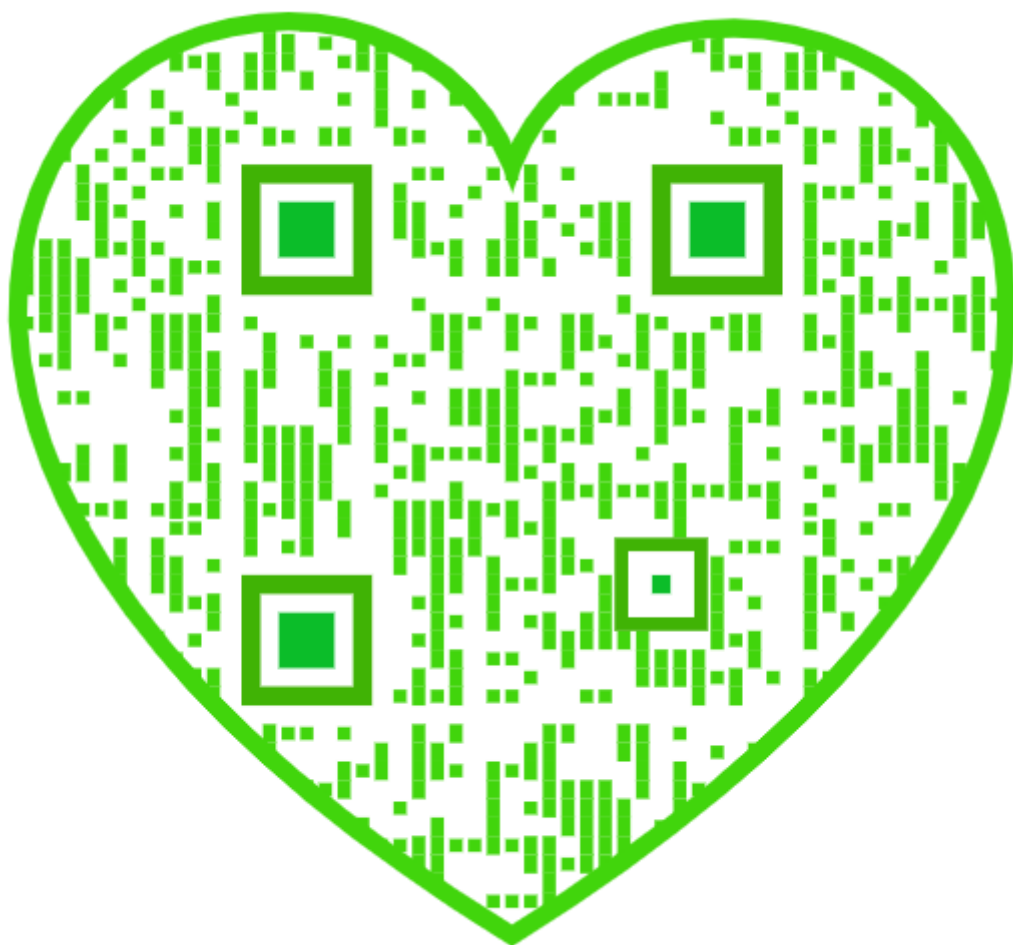
Cataloguing hundreds of proteins in a life-threatening parasite or an organelle, while technically impressive, is no more than frustratingly tantalizing if some understanding of their activities is not also developed. As submitted papers in proteomics grow in number, *Nature* intends to play its part by insisting on conceptual insights from among the great quantities of information that such work will certainly deliver. Researchers and funding agencies need to beware of hype, but should be conscious of the great potential in this research, and keep themselves abreast of the key techniques and technologies. ■

Gene therapy for the public

The hearings held by the US National Institutes of Health (NIH) last week into problems that have arisen in adenovirus-based gene-therapy trials revealed clear breaches of protocol by some of the field's leading researchers, and inadequacies in the state of experimentation (see page 707). Adenoviruses are one of several types of vectors being tried in clinical gene-therapy studies. The case of Jesse Gelsinger, whose death during a gene-therapy trial led to last week's hearing, reveals how poorly understood are the body's responses to those vectors in particular. But the uncertainties and the violations revealed by the hearing should not halt the pursuit of the adenovirus approach, whose particular advantages include the fact that they can infect non-dividing cells.

Given gene therapy's highly experimental and controversial nature, demanding standards of openness are appropriate. Last week's hearings underscore the desirability of giving the NIH's Recombinant DNA Advisory Committee (RAC) more power to examine adverse events as they arise, thus ensuring a public airing of problems.

The NIH has made commendable efforts to make itself accessible — all RAC meetings are open to the public and their minutes are on the web. Given the controversy of the past few months and the worries and even hostilities that experiments in genetic manipulation can engender, the NIH can usefully go further. It should establish itself as a model provider of readable and easily navigable information, providing on its website a substantial overview of the genetic basis of diseases and the state of understanding of respective gene therapies. Considering the amount of unreliable information offered to the public on the web (see page 722), and especially the interests of those members of the public more directly concerned — patients, for example, or their relatives and dependents — the NIH should use the website routinely to give detailed but well-signposted and comprehensible information on the state of trials, covering progress and problems worldwide. Such a programme of enhanced access requires a small budgetary commitment compared with experimental funds. But it could come to play a major role in sustaining both informed consent and public trust. ■





Physics teaching
UK physicists bring
the school syllabus
up to date
p708



Laser target
The US Nova laser
target chamber goes
to France on loan
p709



Mars blow
The loss of the Mars
Polar Lander gives
NASA a headache
p711



X-ray success
European space
scientists relieved
at XMM launch
p713

IBM promises scientists 500-fold leap in supercomputing power...

Paris

Biologists are hailing IBM's US\$100 million project to build a 'petaflop' computer as likely to revolutionize our understanding of cellular biology. The computer, nicknamed 'Blue Gene', would be around 500 times faster than today's most powerful supercomputer.

Computer scientists say that the planned machine, details of which were revealed last week, is the first large leap in computer architecture in decades.

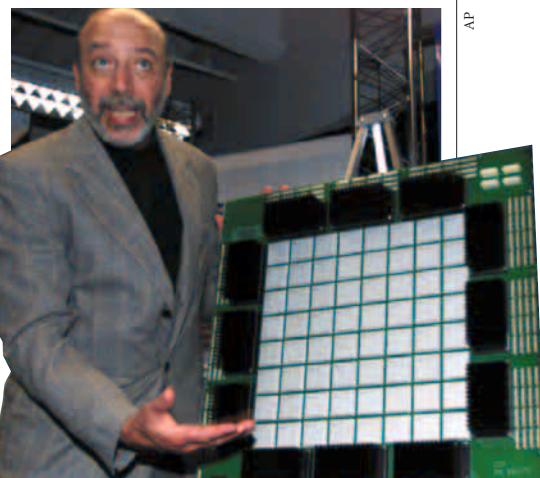
IBM will build the programme around the challenge of modelling protein folding (see below), with much of the research costs going on designing software. It will involve 50 scientists from IBM Research's Deep Computing Institute and Computational Biology Group, and unnamed outside academics.

But Blue Gene's hardware will not be customized to the problem and, if IBM's blueprint works, it will offer all scientific disciplines petaflop computers. These will be

capable of more than one quadrillion floating point operations ('flops') per second — around two million times more powerful than today's top desktops. Most experts have predicted that fundamental technological difficulties would prevent a petaflop computer being built before around 2015 (see *Nature* 402 supp., C67).

"It is fantastic that IBM is doing this," says George Lake, a scientist at the University of Washington and NASA project scientist for high-performance computing in Earth and space science. IBM is showing leadership by ushering in a new generation of supercomputers, he says.

The biggest technological constraints to building a petaflop machine have been latency — increasing the speed with which a chip addresses the memory — and reducing power consumption. A petaflop computer built using conventional chips would consume almost one billion watts of power. IBM reckons Blue Gene will use just one million.



On the board: IBM senior vice-president Paul Horn with a mock-up of a Blue Gene array.

Although processor speeds have increased exponentially, the time to fetch data from the memory of a supercomputer, 300 nanoseconds, is only slightly less than half what it was 20 years ago. Putting more and more transistors on a chip is therefore unlikely to lead to much greater speed.

"We set out from scratch, completely ignoring history, and thought how can we get the highest performance out of silicon," says Monty Denneau, a scientist at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York, who is assistant architect of Blue Gene.

Arvind, a professor of computer science at MIT who is considered one of the top authorities on computer architecture, applauds IBM's approach. "It has made very big steps in rethinking computer architecture to try to do without the components that consume power; it has taken all these research ideas and pulled them together."

One innovation is IBM's use of a prototype technology that combines processor and memory on the same chip. This radically reduces the access time and bandwidth at which the processor can address the memory, as well as reducing power consumption.

Denneau claims that Blue Gene's chips get latency down to just ten nanoseconds. Blue Gene's bandwidth is such that it could

... and a chance to tackle protein structure

Paris

IBM's choice of protein folding as the focus of its project to develop the supercomputer Blue Gene is part of a strategy to position itself as a leader in computational biology, as the post-sequencing phase of the Human Genome Project approaches (see above).

The company has regrouped its life-science activities into a single unit focused on genomics and drug design. Sharon Nunes, senior research manager at IBM's Computational Biology Center at the company's Deep Computing Institute in New York, says that predicting protein structure will be a key use of genome data. "Using today's computers it would take a thousand years to fold a protein," she says.

"Blue Gene is tremendously

exciting," says Christopher Dobson, who works on protein folding at the University of Oxford. "Folding is how the cell converts information from DNA into biological function, how the cell targets proteins to particular parts of the cell, controls and localizes them, and is key to turning activity on and off," he says.

Some are sceptical. Computing power alone is not enough, says Geoff Barton, a scientist at the European Bioinformatics Institute in Cambridge. He argues that solving folding will require enormous progress in understanding the interactions between the atoms in a protein.

"There is no guarantee that if you find the minimum energy conformation it will actually be the

native protein conformation, because the potentials used to model the pairwise interactions are imperfect. There are also difficulties with modelling the true solvent environment of the protein."

Ajay Royyuru, a researcher at IBM's Computational Biology Center, says this will be a major thrust of the Blue Gene project. Dobson says the timing is right in that "many of the fundamental principles of folding are beginning to emerge".

Whatever the reservations about protein folding, few biologists doubt that, if it succeeds, the project's contribution to handling the mountains of data emerging from genome sequencing programmes will be revolutionary.

D.B.

download the entire Internet — around 100 terabytes — in less than one second.

Arvind points out that, although a substantial improvement, this is not enough to allow uninterrupted traffic between the processor and the memory, and bottlenecks in processing would still occur. The conventional solution is to put a special high-speed memory, or cache, in front of the processor, to supply it with the most frequently requested instructions and data. These can then be accessed much faster than instructions and data in the main memory.

The problem facing designers of petaflop computers is that caches consume vast amounts of power. "IBM's real innovation is to have done away with the cache," says Arvind. To do so, it has turned to another prototype technology: multi-threading.

Each of Blue Gene's one million processors can do eight tasks, or threads, simultaneously. If one set of threads is busy, the next instantly takes up the relay. "No-one has attempted multi-threading at this level," says Arvind. "Current supercomputers just don't have this feature."

When built, Blue Gene will have one million processors — each capable of one billion operations per second (one gigaflop) — of which 32 will be placed on each chip. A board containing 64 of these chips would be capable of two teraflops, equivalent to Asci Red, the world's most powerful supercomputer, at Sandia National Laboratory in New Mexico.



Chip ahoy: IBM vice-president Ambuj Goyal.

Asci Red distributes tasks across an array of about 10,000 Pentium Pro chips, and cost almost \$60 million to build. Eight of these boards will be placed in two-metre-high racks, and 64 racks will combine to give one-petaflop performance.

Denneau admits that supercomputers are facing stiff competition from clusters of PCs and workstations, which can deliver supercomputing power at a fraction of the cost (op. cit.). But he says that these are still too slow or unreliable for applications requiring tight coupling and low latency.

The cost of building Blue Gene will be a fraction of that of Asci Red, adds Denneau, thanks to a secret proprietary technology used to etch the silicon in the bulky chips. Most of the costs will be in devising parallel computing models and programming methodologies.

Blue Gene's main drawback, says one scientist, is that it may be relatively short of memory. Arvind says the biggest unknown will be reliability. "To put together a petaflop computer in five years is a very big deal," he says. "It may simply not be possible."

Declan Butler

Germany drags its feet over demand for genome funds

Munich

A widely heralded strategy paper proposing significantly higher public investment in genomics research in Germany appears to have been put on hold. German scientists are concerned that the move could threaten efforts to catch up with other countries.

The paper was drawn up by the research ministry, and had been expected to be made public at the annual meeting of the German Human Genome Project in Munich earlier this month. But Wolf-Michael Catenhusen, secretary of state for research, says the paper will now be presented to the government in early summer next year.

The German research ministry has been giving mixed messages about the cause of the delay, which dashes hopes for more cash for genomics next year. Hans Lehrach of the Max Planck Institute for Molecular Genetics in Berlin, a spokesman for the German Human Genome Project, says the delay could be "disastrous" for Germany's hopes of catching up in genome research after a late start.

The research ministry's strategy paper was drawn up after a rare period of lobbying by the scientific community. This began a year ago at a meeting called by Ernst-Ludwig Winnacker, president of the Deutsche Forschungsgemeinschaft, Germany's grant-giving agency for research in universities.

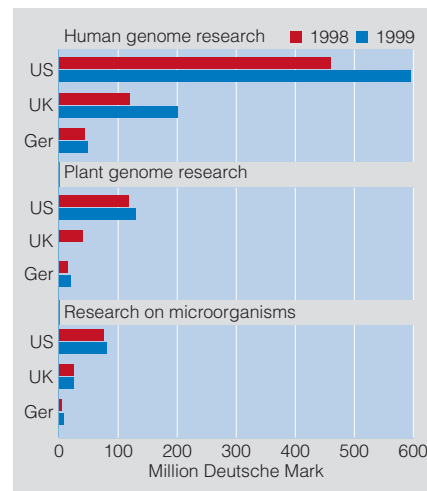
Ministry representatives at that meeting were dismayed by the lack of agreement between scientists on the best way forward for genome research. Various research organizations then prepared individual reports and reached a general agreement that funding should be increased by an order of magnitude if Germany was to compete internationally.

Proposed strategies for spending an enlarged budget were also convergent, focusing primarily on analysis of the function of human genes and development of model organisms for this, but also including some economically, environmentally or medically relevant plant and microorganism genomics.

These aims are reflected in an early draft of the ministry's strategy document which recommends the expansion of the main genome research centres in Munich, Berlin and Heidelberg, and the creation of more focused scientific networks.

According to the draft, such a strategy would require "a new quality of state financing" — it refers to the large differences in the funding of genomic research compared with the United States, United Kingdom and France (see figure). And it should be complemented by industry, and by a reorientation of research institutes' core funding.

Catenhusen, who had been confident in



Public funding of genome research.
(Source: German research ministry.)

the summer that there would be no delay, agrees that the scientific community is now "speaking with one voice", and says there is no dissent in the ministry about the strategy paper's content.

But he now says the paper needs to wait until other decisions are made, such as how the Berlin-based resource centre should be restructured, and how national research centres plan to expand their own genome research through internal reorganization (see *Nature* 402, 450; 1999).

He remains bullish about the strategy paper's big aims. It would be relatively simple to continue modest increases in genome research funding "to a level of say DM90 million [US\$47 million] per year, which would be a 40 per cent increase [since 1996]", he says. "But this is not our strategy's message; genome research funding should be DM250–400 million per year. We need time to prepare ourselves carefully for a round of persuasive discussion with the government."

But many scientists are dismayed by what they see as delaying tactics. Lehrach says: "We lost a lot of time in the last ten years when other countries were moving slowly in genomics, but now genomics is moving very fast, so any months we lose in failing to bolster funds will be a disaster."

While some scientists worry that the delays could reflect splits in the Social Democrat/Green coalition government over a genomics strategy that includes plant genomics, a more likely explanation is Germany's economic squeeze. "If politicians are holding back at the moment, it is because of the unclear financial situation," says one ministry official. "But the will to expand genomics research is very strong."

Alison Abbott

Tighter watch urged on adenoviral vectors...

Washington

Gene-therapy researchers using adenoviral vectors should apply stricter quality controls and more precise monitoring to them because of the 'narrow window' that separates their potential efficacy from toxicity, an advisory panel of the US National Institutes of Health concluded provisionally last week.

The Recombinant DNA Advisory Committee (RAC) met to examine the factors that led to the first death attributed directly to gene therapy (see *Nature* 401, 517; 1999). But the panel also analysed the field as a whole, hearing testimony from a host of researchers who have used the vector in clinical trials.

Many repeatedly mentioned the 'window' and the 'threshold effect' associated with it — that one dose level may cause little or no adverse effect in patients, but a slightly higher dose may trigger apparently disproportionate reactions. Those reactions include immune responses resulting in influenza-like symptoms and tissue damage to the organ in which the vector was administered.

Both of these occurred after Jesse Gelsinger, an 18-year-old Arizona man, received one of the highest doses of adenoviral vector ever given to a human, to treat a partial defect of the gene that encodes ornithine transcarbamylase, an enzyme that removes ammonia from the liver.

Shortly after receiving the vector, Gelsinger developed a high fever. Within the first day, tests showed that he had suffered liver injury and inappropriate blood coagulation. These symptoms had begun to improve when, on the third day, Gelsinger started to have trouble breathing. His vital organs were failing, and his doctors took him off life support on the fourth day.

Inder Verma, a gene-therapy researcher with the Salk Institute in La Jolla, California, and co-chair of the RAC, said he was struck not with the similarities among groups, but with the differences. Some groups measured the strength of their doses in infectious units, while others used particles per kilogram.

"There was really no standard," Verma says, which makes measuring the strength and potential toxicity of one vector against another very difficult. Many gene-therapy researchers also seem unclear about their end points, such as how much messenger RNA ends up in targeted cells after gene therapy.

That end point in the experiment that led to Gelsinger's death was one of many "surprising" findings, according to James Wilson, director of the University of Pennsylvania's Institute for Human Gene Therapy, who oversaw the trial.



Verma: 'no standard' for measuring doses.

Before the autopsy, Wilson expected to find the vector concentrated in the liver, because researchers infused it directly through a catheter into the hepatic artery. But the autopsy revealed "significant" amounts in the spleen, lymph nodes, bone marrow and other tissues.

The autopsy also revealed further abnormalities in Gelsinger's bone marrow, indicating that he may have also been infected with a parvovirus, which, when combined with the adenovirus, may have initiated the chain reaction of adverse events. Finally, when Wilson examined the vector used to carry the gene that codes for ornithine transcarbamylase, he discovered duplicate sequences not engineered in the original.

The RAC concluded that such anomalies underscore the need for more careful measurement at all stages of gene-therapy trials. "Toxicities vary from study to study, vector to vector, patient type to patient type," says RAC member Robert Warren, an oncologist with the University of California at San Francisco.

In its preliminary findings, the RAC advised gene-therapy researchers to take more safety measures, including checking the sequence integrity of a vector before administration, monitoring cytokine levels more closely before and after gene therapy begins, and ensuring that patients have no additional virus lurking in their systems.

Researchers administering adenoviral vectors directly into organs should also consider smaller increases between doses, so that there is less danger when they approach the toxicity threshold.

Paul Smaglik

... with proposal to report all 'adverse events'

Washington

The Recombinant DNA Advisory Committee (RAC) of the National Institutes of Health has been urged to recommend that gene-therapy researchers be required to report all 'adverse events' occurring during their trials, whether or not they are related to the treatment the patients are receiving.

The proposal came during a three-day meeting that discussed the death earlier this year of Jesse Gelsinger (see above). Committee officials struggled to define how much information gene-therapy researchers must submit, both publicly and privately, and admitted that those involved in the Gelsinger trial failed on both fronts.

On the public front, the researchers did not report to the RAC a change in the clinical trial's protocol. The committee had approved the protocol to treat ornithine transcarbamylase deficiency with an adenoviral vector administered intravenously.

But when the RAC was temporarily disbanded in 1996, the US Food and Drug Administration (FDA), with the researchers' cooperation, changed the route of administration. The vector was now to be administered in an artery leading directly to the liver in an effort to minimize the

problem of the vector's wide distribution throughout the body.

"We recognize that we probably should have come to the RAC" after the protocol change, says Mark Batshaw, chief academic officer at Children's National Medical Center in Washington. By the time the trial was under way in 1997, a revived but weakened RAC had lost its authority to approve most gene-therapy clinical trials.

Kathryn Zoon, director of the FDA's Center for Biologics Evaluation and Research, presented "preliminary evidence of approximately two protocol violations" by the researchers. Both involved failure to disclose information to the FDA.

First, the researchers decided to treat Gelsinger despite levels of ammonia in his liver that were higher than the limits allowed for enrolment.

James Wilson, director of the Institute for Human Gene Therapy at the University of Pennsylvania in Philadelphia, where the trial was conducted, defended the decision, saying that Gelsinger's ammonia level at the time of enrolment was fine, technically meeting the requirement of the written protocol. The researchers also treated Gelsinger to reduce the

level before he received the vector.

Wilson also initially denied the second charge — that the researchers should have notified the FDA within two weeks when two patients, each of whom received a smaller dose than Gelsinger, had shown higher than normal levels of liver enzymes.

Wilson pointed out that the group had immediately notified the FDA about two previous patients with similar problems, and agency representatives had allowed the trial to continue. He also said he had filed a written report detailing the liver enzymes eight months before treating Gelsinger.

But Steven Raper, associate professor of surgery at the University of Pennsylvania, accepted the next day that they "should have called the FDA" about each of the patients.

To prevent further such confusion over the reporting rules, the RAC considered a proposal that gene-therapy researchers should publicly disclose all adverse effects that accompany gene therapy.

The proposal goes beyond FDA requirements, as it would require disclosure on both expected and unexpected adverse events in the clinical protocol, regardless of whether they are directly related to the treatment.

P.S.

French research minister waters down CNRS reform

Paris

French science minister Claude Allègre has asked the directors of France's main research agency, the Centre National de la Recherche Scientifique (CNRS), to come up with more moderate changes to the agency's structure than those initially proposed by the government last year.

In a letter to CNRS president Edouard Brézin and director-general Catherine Bréchnac, Allègre asks them to focus on recruiting young scientists and increasing the mobility of researchers. The minister also says the reforms should concentrate on developing new disciplines, reducing administrative loads, and moving towards the "europeanization" of large-scale facilities and scientific evaluation.

To attain these objectives, Allègre writes: "I ask you to concentrate on some essential aspects that permit a deep modernization, without revolutionizing the organism". Earlier, more radical proposals were withdrawn after protests from the scientific community (see *Nature* 397, 463; 1999). Following disputes between research unions and the research ministry, prime minister Lionel Jospin appointed a parliamentary commission to study research reforms.

The commission spent three months consulting French scientists, and reported its proposals to the minister in July. But although much was said about the need to reform the CNRS, no action has been taken within the organization itself.

"This is only the prolongation of a process that has been going on for some time," says Brézin of the new reform proposals. But he is optimistic that it is a step in the right direction. "The project one year ago came as unexpected from the minister and myself, and that was a mistake. This time, I think we have taken a different approach. Now we are instructed to negotiate with the boards and the unions."

The scientific community may be soothed by Allègre's wish to preserve the National Committee for Scientific Research. His desire to slim down the committee was one of the more controversial aspects of last year's proposals. The committee acts as the 'parliament' of the country's scientists, and is involved in evaluating laboratories and administering recruitment.

Brézin says that one of the most urgent needs is to open up the seven scientific departments of the CNRS to allow disciplines to interact.

Heather McCabe

'Relevant' teaching turns UK children back on to physics

London

Twenty-five British schools are piloting a new approach to teaching physics to 16- to 19-year-olds that is already said to be attracting those who might otherwise have deserted the subject.

'Advancing Physics' is intended to bring "physics teaching firmly into the twenty-first century, using relevant and up-to-date examples of physics from everyday life". It is part of a £1 million (US\$1.6 million) initiative by the UK Institute of Physics (IoP) to be launched at next month's annual meeting of the Association for Science Education.

Supported by the Institute of Electrical Engineers and the Particle Physics and Astronomy Research Council, the course uses a range of teaching approaches, examples of applications, and work for students. There is also a CD-ROM with activities that use the computer for doing physics.

The IoP say that schools involved in pilot studies have reported increased enthusiasm for physics among students, and that in one school physics became the most popular choice for sixth-form study. One aspect of the course is the way it uses the manipulation of images to bring together the aspects of what has become a disparate subject.

Mick Brown, a physicist at the Cavendish Laboratory at the University of Cambridge who has been assessing the course for the IoP, describes this feature as "wonderful". "It puts acquisition, manipulation and display of information at the centre of physics, and in doing so unites astronomy, cosmology, high-energy physics and applied physics," he says.

John Ogborn, professor of science education at the University of Sussex, who has been



Turn off? Critics argue that traditional ways of teaching physics need a thorough review.

involved in designing the course, says that the IoP is concerned at the falling numbers of young people going into physics. To make the subject more attractive, Ogborn's group has created what he describes as a package with a core of essential work and an engineering flavour, relating to skills useful in jobs.

But the course's design was constrained by the fact that 60 per cent of the subject content is prescribed by a central authority — the Qualification and Curriculum Authority — which approves A-level courses. Advancing Physics is in the last stages of obtaining official approval.

The core curriculum is intended to let universities know what to expect from students. Ogborn says that those who developed the new course had to be "inventive". The way that sensors work, for example, was used to teach how direct currents operate.

Brown points out that attracting more students to physics will be "a long process that will involve enthusiastic teachers, students and backup on the web".

Natasha Loder

Lots more cash for UK universities

London

The UK government and the Wellcome Trust have announced the biggest investment in the infrastructure of university science for 40 years. Up to £320 million (\$519 million) will go to 45 projects in 27 universities, as part of a second round of awards intended to compensate for years of underfunding (see *Nature* 399, 187; 1999).

Project budgets range from £750,000 to £30 million, although exact amounts will depend on the outcome of a procurement review. Overall, the University of Cambridge looks set to do best from this round of Joint Infrastructure Fund (JIF) awards, with bids totalling £83 million — more than a quarter of the round's fund.

But the largest bid — for more than £30 million — is for a new chemistry research laboratory at the University of Oxford. Successful bids include the refurbishment of the Lovell Radio Telescope at Jodrell Bank (£2.3 million) and a facility for research into dark matter (£3.8 million) at the University of Sheffield.

Michael Dexter, director of the Wellcome Trust, says that many high-quality bids had to be turned down. An award fund of £750 million has received bids for more than £2 billion. The scheme was set up in July 1998, and is funded by £300 million each from the government and the Wellcome Trust, and £150 million from the Higher Education Funding Council for England.

N. L.

Roche's *Taq* patent 'obtained by deceit', rules US court

San Diego

A federal judge in San Francisco ruled last week that the patent for the naturally occurring form of the widely used reagent *Taq* DNA polymerase was fraudulently obtained and is thus invalid.

The decision could mean increased competition leading to cheaper reagent prices, as well as legal action to recover unnecessary costs. And, depending on future court rulings, the patent dispute may also undermine a related patent for the DNA-amplification process of PCR (polymerase chain reaction).

The judge ruled that the US Patent and Trademark Office was misled a decade ago by scientists from the Cetus Corporation, which was bought by the Swiss company Hoffmann-La Roche in 1991 (see *Nature* 397, 460; 1999). Roche purchased the *Taq* patent as part of the deal.

The ruling was issued in a long-running federal lawsuit in which Roche originally accused the Promega Corporation of infringing its patent by selling its own native *Taq* to researchers. This lawsuit began after a disagreement between Roche and Promega over *Taq* licensing terms. Promega, a privately held Wisconsin company, is also involved in challenging Roche's patent in Europe.

After hearing evidence earlier this year during a trial in San Francisco, US district judge Vaughn Walker ruled that there were eight instances in which Roche scientists had engaged in "inequitable conduct" with "intent to mislead" to secure the *Taq* patent.

These involved misstatements or the withholding of scientific information during the patent office's review process. They included withholding details about how *Taq* binds in a phosphocellulose column, misleading statements about *Taq*'s fidelity compared with previously described enzymes, and presenting a key experiment as having been performed when it had not been done.

In his ruling, Walker likened the conduct of Roche scientists to "scientific misconduct". Roche officials deny that their scientists engaged in impropriety, and say they plan to appeal. And they insist their other patents are not affected or threatened.

The judge will hold a hearing next month to begin the process by which Promega will seek financial remedies in the case. If his ruling stands, Roche's patent for native *Taq* will be void and unenforceable.

He ruled that all claims of Roche's patent are unenforceable. The patent includes three claims, one of which is for the native *Taq* and another for recombinant *Taq*. Therefore, Promega argues, Roche's patent covering both native and recombinant *Taq* is unenforceable.

Furthermore, because the judge has ruled that Roche's patent is unenforceable for *Taq*, Promega attorneys plan to argue that Roche's patent on PCR, which relies on the use of *Taq*, is also unenforceable. But Roche disputes this, saying the ruling only affects native *Taq*, which is used by a much smaller proportion of the research community for PCR reactions.

Among the issues expected to come out of the case are *Taq* pricing, the amount Roche is now alleged to have received improperly under its patent, and reimbursement plans.

For instance, *Taq* was sold by some biotech firms for around 16 cents per unit before Cetus was granted its patent in 1989. *Taq* is now sold for at least 30 cents per unit by companies licensed by Roche. These companies pay Roche 15 to 17 cents per unit, claims Promega attorney James Troupis.

The biotech community is speculating on who will try to recover what funds. US government agencies such as the National Institutes of Health (NIH), the National Science Foundation and the Department of Energy are huge purchasers of *Taq*, either directly or through grants to scientists and universities.

Taq is reportedly the most-purchased reagent under NIH grants. Agency officials decline to say what actions might be taken to recover any funds, pointing out that a decision on such legal action would come from the US Department of Justice. Justice officials were unavailable for comment.

Meanwhile, at least one other company, New England Biolabs of Massachusetts, is considering re-entering the market to offer *Taq* (see *Nature* 390, 327; 1997). Its decision, combined with the elimination of Roche's royalty fee, may lead to cheaper pricing for *Taq*. But Roche insists the judge's decision will have little financial impact because the company believes the recombinant *Taq* patent is unaffected.

Rex Dalton



Fair swap? In exchange for the Nova laser target chamber, France will help develop the NIF.

US sends Livermore laser target chamber to France on loan

Washington

The target chamber from the Nova laser experiment at the Lawrence Livermore National Laboratory in California was last week shipped to France. Physicists there plan to use it as an interim component of the Laser Megajoule, a huge laser facility to be built near Bordeaux.

Officials from the US Department of Energy say the chamber is on loan to France as part of a deepening collaboration between nuclear weapons physicists in France and the United States. In exchange for the loan, French scientists will deliver and develop a diagnostic system for the National Ignition Facility (NIF), the laser experimental facility currently under construction at Lawrence Livermore.

Not everyone at Livermore is happy to see the target chamber go. "It's part of a pattern that makes our people a little uncomfortable," says one official there. He points out that the dismantling of Nova and of another laser, the Beamlet, which has been sent to the Sandia National Laboratory in New Mexico, means that researchers at Livermore will have no machines to work on until the NIF is completed in 2003 at the earliest.

Many of the researchers are using the Omega laser at the University of Rochester, New York, a facility comparable in size to Nova. Operating time at Omega has been increased to allow more experiments from researchers displaced from Nova, the Department of Energy says.

Nova cost about \$150 million to build. But energy department officials say that the scrap value of its 5-metre-diameter target chamber — where ten laser beams



► converge to heat a tiny target — was only \$30,000. A senior official values the radiation-temperature diagnostics system, which France will provide at the NIF in exchange for the chamber, at \$1 million.

Unlike the NIF, which was designed to be fully functional by 2003, the Laser Megajoule will be constructed incrementally, starting with a few lasers firing targets in the chamber borrowed from the United States, and later adding lasers and new target chambers to accommodate higher energies.

Both facilities aim to achieve 'ignition', at which the fusion of deuterium and tritium fuel inside the target provides enough heat to sustain itself for a short period.

France and the United States have been working increasingly closely on using lasers for nuclear weapons physics since they signed an agreement to collaborate on this in 1994. It has been reported that the French demanded the agreement in exchange for making a commitment to stop nuclear testing.

According to Matt McKenzie, who monitors nuclear weapons research for the Natural Resources Defense Council, a Washington-based environmental group, the collaboration has moved beyond working together on laser equipment to the design of the targets, called hohlraums, which are heated by the lasers to reproduce conditions inside a detonating nuclear weapon.

Both the Laser Megajoule and the NIF are intended primarily to help train nuclear weapons physicists and to simulate conditions inside nuclear weapons. But they will also be used by plasma physicists to explore the feasibility of inertial confinement fusion as an energy source.

Colin Macilwain

Germany's research agencies open up recruiting practices

Munich

All vacancies for scientific directors at Germany's Max Planck Institutes will in future be advertised, and a worldwide search for new institute heads will be coordinated by central search commissions, announced Hubert Markl, president of the Max Planck Society (MPS), this week.

The reform of the appointment system is one of the moves being taken in response to an external evaluation of Germany's main basic research organizations published last spring (see *Nature* 399, 395–396; 1999).

More than half of the MPS's 240 scientific directors will retire within the next eight years. "Systematic head-hunting" of leading German and international scientists is to begin at least three years before a director retires, according to Markl.

By increasing centralized control over appointments, the MPS is keen significantly to strengthen its influence on the areas of research at the 80 institutes. Adopting a more strategic approach to research funding was one of the principal suggestions of the international evaluation committee.

As for the Deutsche Forschungsgemeinschaft (DFG), Germany's grants agency for university research, the committee complained that its inflexible and conservative funding policy would tend to block promising new ideas.

After six months of intensive discussions, the MPS and DFG have now accepted in principle to put most of the committee's suggestions into action. But officials at both organizations stress that constant annual increases in their budgets — currently around DM2 billion (US\$1.03 billion) each



Markl: reform of recruitment system needed.

— are essential for the proposed reorganization.

Markl and Ernst-Ludwig Winnacker, president of the DFG, presented their organizations' final reports on the evaluation at a press conference this week. They announced plans to create better and more flexible career opportunities for young scientists and women. They also agreed

to cooperate more closely in the training of PhD students, and to promote innovative interdisciplinary research.

But the MPS added the qualification that the number of additional junior independent research groups at its institutes would be limited by the need to ensure an "appropriately high scientific level", as well as by the lack of "sufficient follow-up positions".

The DFG agreed to revise and speed up its peer-review system. It wants to limit the exclusive right of Germany's scientific societies to nominate referees, in the hope of increasing the number of younger scientists and women among elected referees. But it says it will not introduce a quota system.

The committee had recommended that the DFG should "actively influence the long-term developments of science", rather than just responding to them. But the agency argues that the panel "underestimated the DFG's actual strategic activities".

The MPS and DFG are calling on the government to help put the reforms into action by providing adequate finance and relaxing restrictive employment laws.

Quirin Schiermeier

China and Hong Kong pool effort in high-tech projects

Beijing

China's desire to strengthen scientific links with its Hong Kong Special Administrative Region has been confirmed by the announcement of 14 jointly funded research projects.

The Natural Science Foundation of China (NSFC) and the Research Grants Council of Hong Kong (RGC) will finance the research over the next three years. The winning projects were selected from 229 applications by a panel consisting of six experts from the mainland and six from Hong Kong. The work will be funded under a joint research scheme established by the two research councils in November 1998.

Some RMB5 million (US\$604,000) will come from NSFC and HK\$10 million (US\$1.3 million) from RGC each year. The projects are in six fields: new materials, marine and environmental science, life science, management science, information science, and traditional Chinese medicine.

Surprisingly, only one application from the Chinese Academy of Sciences, China's leading research organization, was approved, compared to 11 approved applications from Chinese universities, including the prestigious Beijing University and Tsinghua University.

The NSFC says it has been paying attention to funding scientific exchanges between

mainland China and Hong Kong. About 300 such projects were supported by the foundation during 1991–98.

Before the establishment of the joint research scheme, cooperation between the two sides was sporadic and unorganized, without any security for intellectual property, according to Tang Xifang, an official in charge of the scheme at the NSFC.

"The central government asked mainland China to support the scientific research efforts of Hong Kong, so that Hong Kong can become a centre for high-technology development. The establishment of this grant programme is one step taken in response to the government's call," says Tang.

Tian Xuewen

Have NASA's scientists been the victims of excessive expectations?

The failure of the Mars Polar Lander may have been the result of NASA's high demands. But aiming lower could mean cancelling missions.

Officials at the US space agency NASA were quick to admit after the Mars Polar Lander was pronounced dead last week that the Mars programme is in trouble and needs a complete rethink.

Their hints that the next lander, scheduled for launch in 2001, may be postponed must have had engineers at the Jet Propulsion Laboratory (JPL) sighing with relief. The engineers have known all along that the space agency's plan — launching to Mars every 26 months, and beginning sample collection, a vastly more complicated task, in 2003 — was unrealistic.

But despite occasional warnings, mostly *sotto voce*, that the Mars programme is overburdened and underfunded, top managers at JPL and NASA have been shutting out such negative thoughts (see page 721 in this issue and *Nature* **382**, 481; 1996).

The 1997 Mars Pathfinder landing was taken as proof that the 'better, faster, cheaper' approach works. Failures and near failures in other quarters — such as SOHO, Deep Space 1, Lewis and Clark, and the Wide-field Infrared Explorer — were chalked up to other factors, such as miscommunication or a failure to follow the rules.

The loss of four spacecraft in ten weeks, however, has shaken that faith, and the fallout will extend beyond Mars: NASA's entire science programme is based on optimistic assumptions about how far money, manpower and technology can be stretched.

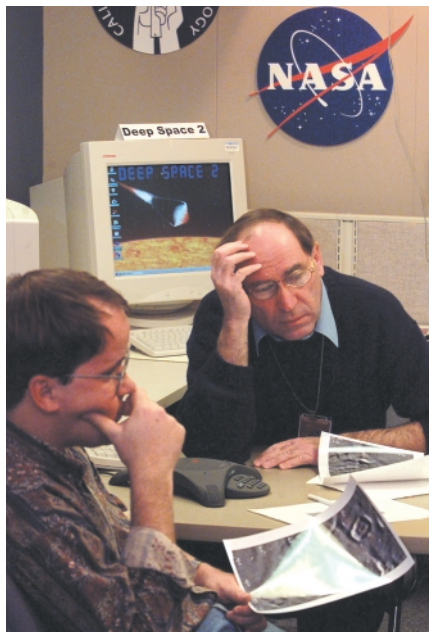
Problems ahead

Responding to the charge that his push for economy has gone too far, NASA administrator Dan Goldin counters that the United States cannot go back to billion-dollar missions or just throw money at the problem.

Most scientists who work with NASA agree. And many agree with Stamatios Krimigis, head of the space department at the Johns Hopkins University Applied Physics Laboratory (APL) in Maryland, who says that 'better, faster, cheaper' is "getting a bum rap".

But David Black, director of the Lunar and Planetary Institute near Houston and a veteran of NASA science advisory committees, says it is dangerous to put "too much emphasis on the 'cheaper' side of the equation".

JPL has had a hard time adapting to the new way of doing business, say critics. They



We have a problem: NASA researchers wait for news of the missing Mars Polar Lander.

say that the lab — which leads the effort to explore Mars and the outer planets, and plays a key role in other large ventures including the Space Interferometry Mission and the Space Infrared Telescope Facility — too often underestimates the complexity of technologically challenging projects, and its managers are too willing to agree to unrealistic demands from NASA.

JPL may be more vulnerable in this regard, as its staff are contractors rather than secure government employees. This means that NASA is a customer, and the customer has to be satisfied.

Other JPL-run programmes are having difficulty as a result, and one outside observer says "the worst is yet to come". Several planetary scientists told *Nature* that planning for the Europa Orbiter mission, which NASA hopes to launch in 2003, and for the Pluto-Kuiper Express the following year, is in serious financial and technical disarray.

Other parts of NASA's science programme have had more success with the 'better, faster, cheaper' approach. Four Discovery planetary missions have made it to the launchpad on time and on budget, although only two — Pathfinder and the

Lunar Prospector — have yet completed their respective missions.

APL's Near Earth Asteroid Rendezvous spacecraft missed its encounter last year but is now on track to reach the asteroid Eros in February. The difference with Discovery missions, says Steven Squyres, a Cornell University planetary scientist who heads NASA's science advisory council, is that they are subject to competitive peer review, which helps to weed out proposals with unrealistic technical or financial assumptions.

Tight deadlines

This is not the case with JPL's Mars programme, however, the requirements for which come down from NASA headquarters. The agency has called for a visit to the planet at every 26-month opportunity, so JPL is designing and building several Mars spacecraft at once, with no possibility of letting the schedule slip, and with far less money and staff than it had when the launch rate was more leisurely.

"The sudden onslaught of many missions has been the challenge," says Tony Spear, who led the successful Pathfinder project before leaving for private industry. He is heading a comprehensive study of 'better, faster, cheaper' missions for NASA headquarters, due next month.

Cancelling the 2001 lander would provide some breathing space in the schedule but may not make sense because much of the spacecraft is already built. JPL managers do not yet know if the Mars Polar Lander failed because of a technical problem or was just unlucky enough to land in an unsafe place.

Squyres, who has been building the scientific payload for the 2001 mission, says that future landers may need a camera to scout out the landing site during descent, along with a way to avoid obstacles if necessary. Asked whether it is possible to visit Mars every 26 months, given the current budget, he admits that he does not know.

NASA needs to ask this question for each science mission. Brad Parkinson, a Stanford University professor who heads the agency's overall advisory council, supports the 'better, faster, cheaper' approach, but concedes that the agency "probably has more on its plate than it can afford". With no budget relief in sight, the inevitable decision will be about which projects to cancel.

Tony Reichhardt

France needs its own synchrotron, says advisory council

Paris The scientific council of France's main research agency, the Centre National de la Recherche Scientifique (CNRS), has insisted on the need for separate national synchrotron sources in France and the United Kingdom. This follows the French government's decision to contribute to a British machine instead of building a third-generation X-ray source in France (see *Nature* **400**, 489; 1999).

The council states that science minister Claude Allègre's decision does not meet the needs of French scientists. It says that French researchers need 21 high-intensity beamlines on straight sections, but that the UK project, Diamond, offers only six. In addition, the possibility of renting out beamlines at other European synchrotrons looks dim, as most are already saturated with users.

The council also finds fault with the decision to effectively scrap France's synchrotron facilities at the Laboratoire de l'Utilisation du Rayonnement Électromagnétique (LURE), at Orsay, close to Paris. This facility hosts 400 scientists and 1,800 annual users.

In a letter to the directors of the CNRS and the French Atomic Energy Commission,

Vincent Courtillot, director of research at the science ministry, calls for discussions to begin on the closure of LURE. He says the centre should be phased out from 2001 and shut down in five years, just in time for Diamond's planned opening.

US cell biologists to go it alone on lobbying

Washington The American Society of Cell Biology is to part company with the Federation of American Societies for Experimental Biology (FASEB), the umbrella organization that lobbies for biomedical funding, on 1 January next year.

Society officials say it is leaving so that it can target fewer, more specific issues, rather than supporting FASEB's broader agenda. "We can act most effectively in the context of a more focused group," says Richard Hynes, the society's incoming president. The society will still work with FASEB when their agendas coincide.

Medical brain drain looms in Australia...

Sydney One in three medical researchers in Australia is considering leaving the field or moving overseas to continue medical research there, according to the first work-

place survey of them. This is despite a boost to funding in the May 1999 budget (see *Nature* **399**, 94; 1999).

In the survey by the Australian Society for Medical Research, the main reasons for dissatisfaction cited by 266 respondents are that salaries in Australia are considerably less than those of Australian researchers working overseas, that job security and opportunities are poor, and that funding of research has decreased over the past five years (the current conservative coalition has been in office for nearly four years).

"More incentive must be given to young people to undertake a career in medical research," says the society's Moira Clay. The surprise was that the recent boost to funding has had little impact, an indication that "the problem runs very deep".

... but Russia's exodus is over, claims minister

Moscow There are only half as many scientists in Russia today as there were nine years ago, admitted the science minister, Mikhail Kirpichnikov, last week. But he said that Russia was not the only country to suffer from a brain drain, and that only 1.6 per cent of those who left science had also left the country.

Speaking in the State Duma, the lower chamber of parliament, Kirpichnikov said

there had been no brain drain in the past two years. One reason was that in 1999 scientists' salaries had passed the Russian average for the first time. The Duma had increased the state budget for science by 2.5 billion rubles (US\$100 million) for 2000.

Challenge to US role in university education

Munich The research ministers of Germany and France, Edelgard Bulmahn and Claude Allègre, have proposed that Europe should set up a joint university ranking system to counterbalance the United States' leading position in university education. According to Bulmahn, a Europe-wide ranking would help students all over the world to compare the pros and cons of the university systems in Europe and the United States.

"Students in non-European countries tend to turn their eyes towards Cambridge alone, but the university landscape of Europe is much more diversified," said Bulmahn. In an interview with the French newspaper *Le Monde*, Allègre recently complained that the private US university system was threatening to foster "standardization of teaching and thought".

A Franco-German working group is to draw up details of the proposed ranking over the next few months.

Consortium will focus on post-genomic software

Paris The Pasteur Institute, the French national computing research agency, INRIA, and two biotechnology companies, Hybrigenics and Genome Express, have announced the creation of a bioinformatics consortium to design a modular software platform for post-genomics studies.

The consortium, called Genostar, will develop informatics systems to analyse genes and proteins and compare interactions between them. The members of the consortium say the applications are destined to go beyond the identification of genes, and will be able to help characterize gene functions and find pharmaceutical targets.

US cancer society to fund stem-cell work

Washington The American Cancer Society (ACS) last week endorsed US government funding of human embryonic stem-cell research, and said it will use its own funds to support research in the area. The move reverses the perception that the huge cancer charity had withdrawn its public support for the controversial research because of pressure from Roman Catholic supporters.

In the summer, the ACS withdrew from a

coalition of disease groups lobbying Congress to fund stem-cell research after lay Catholics protested and a top church official sent a letter of complaint (see *Nature* 400, 493; 1999). The ACS said at the time that it had not yet formulated its position.

"Stem-cell research holds extraordinary promise in the overall effort to eradicate cancer," says Harmon Eyre, the charity's chief medical officer. The ACS stressed that it will only consider research proposals that follow National Institutes of Health guidelines.

Space telescope sends back a self-portrait



London Europe's X-ray satellite was launched successfully on Ariane 5 last week from Kourou in

French Guiana. The 3.9-tonne X-ray Multi-Mirror (XMM) space telescope — the first commercial payload for Ariane 5 — sent back pictures of itself in space only five hours after taking off, using two micro-cameras, at 55,000 kilometres above the Earth.

Scientific objectives for the telescope include finding out what happens in the vicinity of black holes, and increasing our understanding of gamma-ray bursts and cannibalism among the stars.

ESA

MATTER OF EXPRESSION

Cell talk: a brief guide to the vocabulary of proteomics

p715

METHODS MADE SIMPLE

How the latest technologies separate and identify proteins

p716

FAST FORWARD

Chip-based technology may speed the pace of the research effort

p718

FUNDING FLOWS

Round-up of who's spending what across the world

p720

A post-genomic challenge: learning to read patterns of protein synthesis

The new science of proteomics promises exciting insights into the working of the cell. But major technical hurdles remain to be overcome.

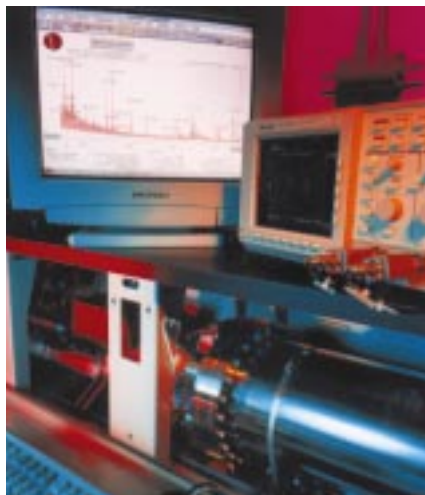
As the full-length sequences of target genomes come into view, biologists know they represent just the start of a long march towards an understanding of how organisms, including humans, develop and function. To many, the next key landmark will be an overview of the characteristics and activity of every protein that an organism can synthesize in its lifetime: its 'proteome'.

Because the technology required is still at a rudimentary level of development, and is more troublesome than genomic technologies, completion of such a task cannot be expected for many years. But that is not stopping biologists from setting out along the way.

Ten years ago, a protein chemist would have been happy to identify two or three proteins a year. Now, increasing volumes of genome data, combined with mass spectrometry technologies, permit a researcher to identify hundreds of proteins within a week.

Although the technologies that have revolutionized protein chemistry are in their infancy, some scientists are already starting to cautiously discuss the concept of a Human Proteome Project, to complement the Human Genome Project. Others say it is far too early to speak so ambitiously, or suggest that the idea of adopting the systematic approach of the genome project is not sensible, given the ever-shifting populations of proteins.

But the idea is seductive. By identifying proteins on the scale of the proteome —



Fast fingerprinting: highly automated mass spectrometry identifies proteins rapidly.

which can involve tens or even hundreds of thousands of proteins, depending on the state of the cells being analysed — proteomics can answer fundamental questions about biological mechanisms at a much faster pace than the single-protein approach.

The 'global picture' painted by proteomics can, for example, allow cell biologists to start building a complex map of cell function by discovering how changes in one signalling pathway — the cascade of molecular events sparked by a signal such as a hor-

mone or neurotransmitter — affect other pathways, or how proteins within one signalling pathway interact with each other. The 'global picture' also allows medical researchers to look at the multiplicity of factors involved in diseases, very few of which are caused by a single gene.

Such understanding can be exploited to identify better drug targets — proteins with which drugs selectively interact to achieve a defined therapeutic response. It also offers drug companies the prospect of being able to rapidly identify the ultimate number of useful protein drug targets — reckoned to lie between 1,000 and 4,000 — and to identify individuals destined by their proteomes to suffer side effects from drug therapy.

However, major technical problems need to be resolved before proteomics can become the large-scale, highly automated affair that genomics has become. There are two principal steps in proteomics (see page 716). The first is separating proteins in a sample, usually using two-dimensional (2D) gels which separate according to molecular charge in one direction and mass in a perpendicular direction. The second is identifying the separated proteins, typically using new mass spectrometry techniques and bioinformatics. Each step has limitations that make automation difficult, but the strongest curses of proteomics researchers are reserved for the 20-year-old 2D gel

Proteomics, transcriptomics: what's in a name?

The term **proteome**, coined in 1994 as a linguistic equivalent to the concept of **genome**, is used to describe the complete set of proteins that is expressed, and modified following expression, by the entire genome in the lifetime of a cell. It is also used in a less universal sense to describe the complement of proteins expressed by a cell at any one time.

Proteomics refers to the study of the

proteome using technologies of large-scale protein separation and identification. The nomenclature is catching on. The generation of messenger RNA expression profiles is referred to as **transcriptomics**, as these are based around the process of transcription. And the complement of mRNAs transcribed from a cell's genome is called the **transcriptome**.

The company CIPHERgen, based in Palo

Alto, California, is trying to popularize the term **phenomics** to describe the technology of automated functional analysis of proteins. The word derives from **phenotype** — the observable characteristics conferred by a gene. Some enthusiastic researchers in the field are even starting to refer to the whole operation of molecular analysis of a cell, extending from DNA through RNA to protein, as **operomics**.

► separation technology. Only the truly green-fingered can reproducibly separate a complex solution of proteins on a gel matrix.

Despite much refinement, it is still referred to as 'the black art' and 'the dark force of proteomics' — but 'a necessary evil', which requires skilled staff. There is a limit to how many proteins a single gel can separate, and the sensitivity is still not adequate to detect proteins appearing at very low levels. Some classes of protein, particularly hydrophobic membrane-bound proteins that are favourite targets for drug development, will not run on 2D gels. On the other hand, the 2D gel is the only method that can resolve large numbers of proteins in a quantitative way.

After the genome

Now that the sequences of genomes from several species have been, or are about to be, completed, researchers and their funding bodies — both public and private — are looking towards the next step: the understanding of gene function. Of the roughly 140,000 genes in the human genome sequence, which will be completed within four years, the function of an estimated two thirds is likely to be unknown.

Protagonists argue that proteomics is one of the most important of the so-called 'post-genomic' approaches to understanding gene function because it is the proteins expressed by genes that are ultimately responsible for all processes that take place within the cell. But, while proteins may yield the most important clues to cellular function, they are also the most difficult of the cell's components to detect on a large scale.

A second, complementary, post-genomic approach is expression profiling, also known

as transcriptomics. When a gene is expressed in a cell, its code is first transcribed to an intermediary 'messenger RNA' (mRNA) which is then translated into a protein. Transcriptomics involves identifying the mRNAs expressed by the genome at a given time. This gives a snapshot of the genome's plans for protein synthesis under the cellular conditions at that moment.

Transcriptomics has the advantage over proteomics that the technology is simple and lends itself readily to automation and high throughput. Messenger RNAs can be fished out of a cellular soup onto microarrays. These are chips onto which are stuck thousands of complementary DNAs derived from as many genes in a genome as have been sequenced, or are deemed necessary for a particular experiment. Each cDNA specifically binds the mRNA transcribed from the gene from which the cDNA is derived. If only a small amount of mRNA is present, this can be 'amplified' by polymerase chain reaction techniques.

Transcriptomics can therefore yield important biological information about what genes are turned on, and when. But it has the disadvantage that, although the snapshot it provides reflects the genome's plans for protein synthesis, it does not represent the realization of those plans.

The correlation between mRNA and protein levels is poor, generally lower than 0.5, because the rates of degradation of individual mRNAs and proteins differ, and because many proteins are modified after they have been translated, so that one mRNA can give rise to more than one protein. Even in the simplest self-replicating organism, *Mycoplasma genitalium*, there are 24 per cent more proteins than genes, and in humans there

could be at least three times more. Post-translational modification of proteins is important for biological processes, particularly in the propagation of cellular signals, where, for example, the attachment of a phosphate group to a protein can trigger either activation or inactivation of a signalling cascade.

So measuring proteins directly gives a more accurate picture of a cell's biology and, if it were as technically easy as measuring mRNAs, more scientists would choose to do so. Ian Humphery-Smith is one scientist who believes that technical difficulties should not stand in the way of a systematic effort to analyse the whole human proteome. His former group, from the Centre for Proteomic Research and Gene Product Mapping at the University of Sydney, has published the most complete proteome to date — almost three-quarters of the predicted proteome of *M. genitalium*, whose relatively small genome (less than 500 genes) has been fully sequenced.

"It's small, and it is not complete, but three quarters of everything is very relevant", says Humphery-Smith, who has recently taken up the new chair of pharmaceutical proteomics at the University of Utrecht in the Netherlands. Proteomic researchers are at the same scientific and technological stage as genome researchers were in 1986 when the Human Genome Project was first seriously proposed, he says, "so the time is right to launch a Human Proteome Project now".

This project should identify, catalogue and annotate the entire human proteome — the expression of every gene: "Now we can do simple prokaryotes, we should start systematically to analyse the proteomes of lower eukaryotes whose genomes are near completion, so that we will be ready to approach the

How to spot a protein in a crowd

The two key steps in classical proteomics are the separation of proteins in a sample derived from cells or tissues, and their subsequent identification. The best separation method is 2D gel electrophoresis, in which spots of a carefully prepared mixture of proteins extracted from cells or tissues are applied to a polyacrylamide gel.

The proteins, which can number tens of thousands, are separated along the gel in one direction according to their molecular charge, by applying an electric field. This process is called isoelectric focusing.

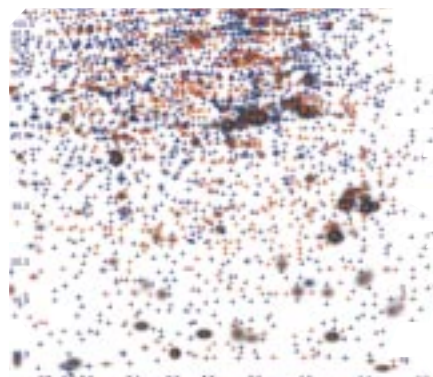
They are then separated in another direction on the basis of their molecular mass, using SDS-PAGE (polyacrylamide gel) electrophoresis, a molecular sieving method. A typical gel can reliably separate 2,000 protein spots in this way. The 'best' gels can separate up to 11,000 protein spots.

Proteins separated on the gel are stained

using the Coomassie blue dye, silver stains, fluorescent dyes or by radiolabelling, and then quantified using spectroscopic or radiographic techniques. As yet, amounts of proteins of less than one nanogram cannot be routinely measured.

Although 2D gel electrophoresis gives the highest resolution of all available methods, it is notoriously difficult to carry out, and it cannot detect some of the most interesting proteins in a cell.

Hydrophobic proteins, for example, such as the cell-membrane-spanning receptor proteins that are such attractive drug targets, simply do not dissolve in solvents used for isoelectric focusing. Neither do proteins with very high relative molecular mass. Low-abundance proteins, which tend to do the most interesting jobs in a cell, are drowned out by high-abundance 'housekeeping' proteins, which can be



Separated out: a 2D gel displays more than 2,000 proteins, known (red) and unknown (blue).

present at 10,000 times the concentration.

Furthermore, there is no amplification method for proteins, analogous to the

DANISH CENTRE FOR HUMAN GENOME RESEARCH

human proteome when the genome is completed in 2003. This would allow essential infrastructure to be put in place that could also be used for the human proteome."

Mathias Uhlén, director of the new Centre for Proteomics at the Royal Institute of Technology in Stockholm, also believes it is "time to talk about a global proteomics project — and to set endpoints for the project which obviously cannot be as simple as the single endpoint of the Human Genome Project".

And Hans Lehrach, a director of the Max Planck Institute for Molecular Genetics in Berlin, and spokesman for the German Human Genome Project, sees the systematic cataloguing of the proteome as almost a duty for the genomics community, to provide full information for biological researchers.

This view is not widely shared among the as yet small international community of proteomic researchers. Most feel that the technical hurdles are still too high to launch such a major initiative in the near future. But Humphrey-Smith argues that serious investment in a Human Proteome Project would rapidly drive the development of technologies to complement the limited 2D gel approach.

Craig Venter, the genome technology guru whose company Celera Genomics aims to sequence the human genome by 2001, is convinced that this is the right way for scientists to work. "Scientists must work with the tools at hand," he says. "It would have been wrong for molecular biologists to have waited until the sort of high-throughput technology we now have at Celera became available before launching the Human Genome Project."

But Leigh Anderson, director of one of the world's two largest proteomics companies, Large Scale Biology Corp., which works

Proteomics is one of the most important of the 'post-genomic' approaches to understanding gene function.

with most of the world's top pharmaceutical companies on specific proteomics projects, is one scientist who is unenthusiastic about a Human Proteome Project. He actually proposed such a project, then called the Human Protein Index Project, nearly 20 years ago, which was discussed at high political levels before being superseded by the genome revolution. "But we were naïve then," he says.

He now thinks that "large-scale proteome research should be carried out in companies rather than through government-sponsored projects". Unlike the genome, which is a constant entity, interesting biological and medical questions relate to the proteome expressed in different tissues, different disease states and under different conditions, he says: "This prospect of immediate application is why proteomics is so commercially interesting at the moment".

"Enthusiasm, scepticism and indifference: this is the same range of opinions that were expressed when the Human Genome Project was launched," says Mike Ashburner, co-director of the European Bioinformatics Institute at the Sanger Centre in Cambridge, England. Scientists were daunted by the size

of the task of sequencing genomes from whole organisms, he says, and many questioned whether the investment required was justified. But experience bears out the wisdom of taking the plunge.

Clinical applications

While some scientists talk in terms of ambitious 'cataloguing' plans, pharmaceutical companies are already working on the application of proteomics in disease, hoping to identify new diagnostic tools or leads for the development of drugs.

Identifying disease markers, proteins that appear or disappear during the course of a disease, does not necessarily require that all expressed proteins in a clinical sample be identified — although the more complete the proteome, the more complete will be any set of markers. Service companies such as Large Scale Biology Corp. and its UK competitor Oxford GlycoSciences (OGS), which have developed automated and high-throughput proteomics technologies, can identify high proportions of the total protein expressed in a tissue with reasonable efficiency.

For example, under a contract with Pfizer, OGS has analysed hundreds of samples of cerebrospinal fluid from patients in different stages of Alzheimer's disease, collected by Pfizer and the National Institute of Mental Health, a division of the US National Institutes of Health (NIH). Results have not yet been formally announced. But key proteins have been identified whose expression changes during the progression of the disease.

Scientists in such big companies tend to be scathing about the attempts of small academic groups to set up similar clinical research projects with more modest

Current methods separate proteins on gel, then identify them using lasers.

polymerase chain reaction method for amplifying genes, to help make up for this difference in abundance. And proteins with very high molecular charges or very low molecular mass will not separate on gels.

More proteins can be separated by using time-consuming tricks such as fractionating cells into their various organelles before running consecutive gels, or sequentially running 'zoom' gels, which separate proteins according to a series of narrow isoelectric ranges. But, even after such tricks have been exhausted, around 20 per cent of the proteome may still be missing.

Once proteins have been separated, visualized and quantified, they must be identified. Spots are excised from the gel — companies such as Oxford GlycoSciences and Large Scale Biology have robots to do this — and proteins are digested into fragments by specific proteases such as

trypsin. The fragments are analysed by mass spectroscopy in a process called peptide mass fingerprinting, in which proteins are identified by comparing the mass of the peptide fragments with data predicted by genetic or protein sequence information.

The most commonly used type of mass spectrometer is called MALDI-TOF, in which protein fragments in a solid-phase sample are ionized by a laser beam. The MALDI-TOF is highly automated, and popular with academic research groups.

A more sophisticated, but more time-consuming, method is tandem mass spectrometry. Each peptide analysed by mass spectrometry is subjected to further fragmentation and mass spectrometry, to give partial information about the peptide sequence. This information is helpful if full genomic data are missing, because proteins can be identified from databases of

expressed sequence tags, which have more entries than standard genomic databases.

Any experiment that involves a limited number of proteins, such as those in a purified protein complex, can avoid using 2D gel separation, which is tedious and difficult, by using methods such as high-performance liquid chromatography, gel filtration or one-dimensional gel chromatography.

These methods separate and concentrate proteins in a complex sufficiently well to be identified by mass spectrometry, although they do not allow quantification. The ultimate goal, of course, is to be able to avoid having to separate proteins at all. The protein sample would merely be run over arrays impregnated with antibodies, or other types of affinity probe, to identify all proteins of interest — this is the promise of chip technology (see page 718).

A. A.

equipment. Chris Ashton, until recently with OGS, doubts that small units without access to sophisticated robotics to allow high throughput should get involved. "Proteomics is a big science which is technically much more complicated than genomics, and so best handled by big companies or their equivalent," he says.

But Peter Mose Larsen, director of the Centre for Proteome Analysis at Odense University in Denmark, is one of many academics who disagree. His lab has identified and patented two groups of protein markers of hypertension which are being tested by five pharmaceutical companies as diagnostic tools, and in some cases as drug targets.

He admits that inexperienced groups often underestimate the skill required in running 2D gels. But lots of small-scale science is required, which does not need the highest proportion of expressed proteins to be detected, nor the most efficient robotics, he says. "The contribution of many small groups is vital; and academic groups can work as well as large companies, with equal success." His centre is one of ten partners in a European Commission initiative to develop generic proteomics technologies and make them available to smaller laboratories, and also small companies, throughout Europe.

In fact, small — and large — academic research units springing up around the world are starting to identify markers in a range of conditions, from organ rejection to cancers, and a wave of publications is likely to hit the literature soon.

Two centres of the Ludwig Institute for Cancer Research, for example, in London and Melbourne, are running programmes to determine protein expression in breast and colon cancers. Director Mike Waterfield says

these programmes, carried out in collaboration with OGS, "will complement the transcriptomics and genomics efforts that the Ludwig Institute is carrying out with the Imperial Cancer Research Fund and the Sanger Centre near Cambridge". This collating of information at all three molecular levels, sometimes jokingly referred to as 'operomics', is an increasingly frequent feature of research programmes.

The National Cancer Institute (NCI), a division of the NIH, funds similar research through an extramural programme concerned with developing and using molecular-based technologies, including proteomics and transcriptomics, to aid tumour classification.

One of the biggest centres to be funded is that of Samir Hanash, from the University of Michigan School of Medicine, who has a new \$10 million NCI grant to generate proteomic data on lung, colon, liver, breast and ovary cancers, which will complement genomic and transcriptomic data from the same tissues, and create a database combining all the information, including original 2D gel images.

The NCI is also co-sponsoring, with the Food and Drug Administration (FDA), a new multi-million dollar Tissue Proteomics Initiative run by Lance Liotta, a principal investigator at the NCI laboratories in Bethesda, Maryland, and molecular biologist Emanuel Petricoin, a senior staff fellow at the FDA. This will generate proteomic data in a range of cancers at different stages of disease and therapy — as well as developing proteomic technologies suitable for clinical work.

The FDA is interested in particular in finding surrogate markers of toxicity and measures of the efficacy of both new and existing drugs. In one project, Petricoin is



One among many: the p-21 H-ras protein, which is implicated in cell-signalling functions.

monitoring changes in protein profiles, and some signal transduction pathways, in breast cancer epithelium and in body fluids before and after treatment with herceptin, a humanized monoclonal antibody that was recently approved by the FDA for late-stage breast cancer treatment. The changes will be correlated with side effects and clinical efficacy.

Petricoin says the programme aims to demonstrate 'proof of concept' to drug companies that remain uncertain about the impact proteomics is likely to have on their drug development programmes.

Petricoin is convinced the impact will be large, allowing, for example, companies to predict the toxicity of a potential drug much earlier than at present by looking for surrogate markers in the tissues of healthy volunteers in phase one clinical trials. At present, and in the absence of surrogates, toxicity often only becomes apparent when a drug is used in

Can researchers find recipe for proteins and chips?

Automated chip-based technologies for analysing thousands of proteins simultaneously, analogous to the cDNA chip-based technologies that have facilitated transcriptomics, could provide a leap forward for proteomics research, whose progress is limited by the cumbersome multi-step methods currently available.

Since the chemistry of proteins is orders of magnitude more complicated than DNA chemistry, the leap forward is unlikely to be seen in the next few years. Despite this, scientists are already starting to work on the problem, using different approaches.

The Stanford University laboratory of Patrick Brown, for example, renowned for its success in developing cDNA microarrays for high-throughput analysis of gene expression, is investigating whether similar success can be achieved with proteomics.



Brown: success on microarrays.

Neither Brown, a paediatrician with a doctorate in biochemistry, nor his postdoc, chemist Brian Haab who is working on the project, are daunted by the difficulty of the task. "Everyone recognizes it is worth doing," says Brown.

Haab hopes to become the next young star from Brown's lab, which has already produced Dari Shalon, inventor of a

robotic arm that is an important component of microarray analysing systems, and Joseph DeRisi, famous for producing elegant experimental results of gene expressions.

Funded partly by the US National Cancer Institute, Brown's proteomics project has

completed a pilot study designed to demonstrate the feasibility of microarray analysis of proteins. Antibodies are attached to microarrays, and these bind to and identify proteins that have been tagged with fluorescent dyes to allow quantification. Haab has already been able to correctly identify 40 known proteins.

Researchers in Brown's laboratory hope to compare hundreds or thousands of protein expressions to identify a pattern of variation from human sera or fluid samples, as is now being done with 2D gel technology — but much faster.

Some remain sceptical whether the microarray concept will work broadly enough on proteins, whose diverse chemistry poses serious difficulties, to be helpful in large-scale expression proteomics where thousands of unknown proteins must

large patient cohorts in phase three clinical trials, when tens of millions of dollars may already have been spent unnecessarily.

Also, he says, drug companies could provide information on specific protein markers of toxicity or responsiveness to clinicians who could then screen out patients likely to be sensitive to toxic effects, or unresponsive to a drug, and also monitor patients for early signs of toxicity and efficacy.

Petricoin says that such screening could be done without using 2D gels. Technologies such as SELDI and liquid chromatography and mass spectrometry (see page 716), which are easier to automate, would suffice because the relevant proteins would be known, and would be limited.

Basic biology

These simpler proteomics techniques, which sidestep the 2D gel problem, are also being used by basic biologists to address fundamental questions, such as how proteins in a cell interact with each other in response to internal and external signals. This approach represents a step in the direction of creating a 'virtual cell' which will allow biologists to approach the ultimate question: how do living organisms operate as systems at the molecular level? (See also *Nature* 402, 219; 1999.) The number of proteins involved in a typical pathway of protein interactions is usually less than 100.

Signalling pathways, the chain of protein interactions and other molecular events that follow activation of a receptor protein by a hormone or neurotransmitter, are a fundamental part of this activity, and a major new focus of research. For example, Tony Pawson, head of the programme in

molecular biology and cancer at the Samuel Lunenfeld Research Institute at the Mount Sinai Hospital in Toronto, recently received a Can\$12 million (US\$8 million) grant to study signalling networks on a large scale.

Another key player is Matthias Mann at the University of Southern Denmark in Odense. Mann recently identified the 70-odd components of the spliceosome, a complex of proteins in all cells that act together to ensure that non-coding regions are cut out of mRNA molecules before these mRNA proteins are translated into proteins. Mann pioneered the 'affinity probe' approach to analysing protein interactions. This involves taking one protein, or a part of a protein that has an important binding domain, and simply seeing what sticks to it.

Another approach to analysing protein interactions is the so-called yeast two-hybrid approach, developed by Stan Fields at the University of Washington in Seattle, which uses the yeast genome as a sort of matrix for studying interactions between two proteins which can be from any species.

Hybrids are made of two proteins whose potential interaction is to be analysed: one test protein is attached to a yeast DNA-binding protein, and a second to the so-called transac-

tivation domain, or gene activator, of a yeast transcription factor. If the two test proteins interact, a 'reporter' gene, typically a gene that turns yeast blue, is activated by the transactivation factor which has been pulled into the gene's vicinity on the yeast genome by the DNA-binding protein. This is a potentially powerful technique for large-scale studies.

Pawson says that biologists, who have traditionally devoted their careers to the study of one protein in one pathway to broaden their perspective, are now realizing that more can be gained by using the new technologies to look at all proteins in a pathway at once. It is important, he says, that biologists work with scientists from other disciplines, such as informatics and engineering, to ensure that the right technology is developed. "Like genomics, proteomics biology is becoming technology-led."

Pawson is the first to agree, however, that technology is necessary but not sufficient for the advancement of proteomics. Julio Celis, director of the Danish Centre for Human Genome Research, has a word of caution for starry-eyed newcomers to proteomics: "New technologies are not a substitute for scientific rigour." Celis is often called the founding father of proteomics because he was the first to develop methods to identify multiple proteins separated by 2D gels in the early 1980s.

"Whoever manages to master the biology of the system they are working on will dominate the proteomics race," he says. Just seeing the level of expression of a protein rise or fall in a tissue sample may not be biologically meaningful. "You need to know if the protein in question has been transferred to a different cellular compartment, if it has disappeared from one cell type only to appear in a neighbouring cell type, or whatever."

The prospect of immediate application is why proteomics is so commercially interesting now.

Chip-based technology could speed the pace of protein identification.



Aebersold: 'take a global view'.

be identified. DNA is static, the protein complement of a cell is constantly changing, and proteins fold in many known and unknown ways that affect their function.

"It will be very difficult to find binders, or antibodies, for every single protein, and then detect what is bound to each in a quantitative way," says Ruedi Aebersold of the department of molecular biotechnology at the University of Washington in Seattle.

"If this is to be a global effort, you have to have conditions that will work for all proteins," says Aebersold. "There are more straightforward ways to deal with proteins than what they are doing in the Brown lab."

But Aebersold agrees with the Brown

lab's approach to using protein arrays for disease diagnosis. "That type of assay has a lot of potential in diagnostics, where you look for a particular constellation of proteins. You optimize conditions for specific proteins; you know what you are looking for."

Aebersold's own lab is developing a different 'chip' approach, which involves miniaturizing classical electrophoresis technologies in a way that is compatible with automation. Proteins are separated on channels scratched onto chips by applying an electric field. They are then sprayed directly from the chip into a mass spectrometer for identification.

According to Ian Humphery-Smith, of the University of Utrecht in the Netherlands, both methods suffer the intrinsic problem of miniaturization — the small volume of

protein solution that can be applied to the surface of a chip, which limits the detection of low levels of proteins. Detection of low-abundance proteins is a problem inherent to all proteomics technologies, but it is exaggerated as analysis systems are reduced in size. "We will need a major technological advance in protein detection systems for chip technology to move forward."

A second "absolutely necessary" technological advance required for the Brown approach, he says, is in the area of antibody production, which is slow and cumbersome. Scientists are already tackling this problem (see page 720).

It may take many years. But if a Human Proteome Project is to be achieved, then array-based procedures of one sort or another will be needed to cope with the enormity of the task.

Rex Dalton & A. A.

His own clinical research programme in bladder cancer not only generates classical mRNA expression and proteomic data from normal and diseased biopsy samples, but also identifies the exact cellular location of proteins in the same biopsies using specific antibodies. This is time-consuming, particularly as most of the antibodies required have to be made in-house. Classical production of monoclonal antibodies in mice takes many weeks.

Many biologists believe that finding new ways to make high-quality antibodies on a large scale could lead to a second proteomics revolution. Promising ideas for efficient antibody generation are already being tested. One of these, pioneered by Greg Winter at the Laboratory of Molecular Biology in Cambridge and the company Cambridge Antibody Technologies, uses bacterial phages to generate antibodies. Another idea, pioneered by the Swedish Centre for Proteomics in Stockholm, uses combinatorial protein chemistry to generate potential artificial antibodies.

If these, or any other method, could lead rapidly to the creation of a library of highly specific, high-affinity antibodies that could

reliably identify each individual protein in the proteome, proteomics could become a highly automated, high-throughput science like genomics. The antibodies could eventually be placed on chips, much like the cDNA arrays used in transcriptomics (see page 718).

Even in the near future, however, proteomics conducted on a limited scale has much to offer clinicians and basic biologists, and the decisions of funding agencies show that this is becoming more widely recognized (see below). Indeed, these decisions suggest that it may not be premature to start to discuss the realization of a longer-term goal, a Human Proteome Project.

Such a move would provide an invaluable tool for basic biologists and pharmaceutical companies alike, just as the Human Genome Project is providing a freely available tool in the form of complete genomic information. It would be several orders of magnitude more complex. But it would help drive the development of technologies to speed up the task, leading to a far more complete understanding of the operation of the human body than genomics alone can ever provide. **Alison Abbott**

Germany has not thrown itself into proteomics to a great extent, although ironically 2D gel technology was invented in 1975 by Joachim Klose, now at the Max Planck Institute for Molecular Genetics in Berlin. Last year the federal research ministry awarded DM14 million (US\$7.3 million) for five years to set up a proteomics centre in Rostock to improve techniques and evaluate them in clinical projects in east Germany.

But this is a fairly isolated action. No dedicated proteomics project will be funded in the second round of the federal government funded German Human Genome Programme, even though this has prioritized post-genomic technologies.

Peter Jungblut, protein-analysis group leader at the new Max Planck Institute for Infection Biology in Berlin, uses proteomics approaches to define new targets for vaccine development. He says the attitude in Germany is "disappointing — but it is typical that Germany waits for new ideas to be developed in other countries before catching on".

Japan, like Germany, is hesitating before embracing large-scale proteomics. The Science and Technology Agency is requesting a ¥284 million (US\$2.7 million) investment in proteomics research from the government.

"But the government is reluctant to make a full commitment to proteomics," says Teruhisa Noguchi, former director of the Helix Research Institute, a genomics centre that is contemplating a shift in emphasis towards proteomics, whether or not major government investment comes through. Helix is funded by the Ministry of International Trade and Industry and ten pharmaceutical companies.

Switzerland has no funding specifically earmarked for proteomics. But the Geneva-



Radda: expects many grant applications.

based Swiss Institute for Bioinformatics is a proteomics flagship for the country, looking after several public protein databases. Headed by Denis Hochstrasser, the institute is financed by the city of Geneva, and is run on proceeds from a new commercial arm, GeneBio, which licenses pharmaceutical companies access to its protein databases (see *Nature* **394**, 214; 1998). GeneBio returns 75 per cent of its profits to the institute.

The pharmaceutical industry is watching developments with interest. Most companies have modest proteomics programmes in-house, and often collaborate with academic groups on projects likely to drive the technology. Many use service companies, such as Oxford GlycoSciences and Large Scale Biology, for larger clinical projects. But, before investing in large in-house programmes, they are waiting to see if proteomics fulfils its potential as a discovery tool. ■

Funding agencies move hesitantly to embrace novel approaches

One of the first funding agencies to recognize the potential of protein analysis was the US National Science Foundation. In 1989 it agreed to support the start of a ten-year programme at the University of Washington in Seattle to create a centre in molecular biotechnology, specializing in the development of proteomics tools.

Others were relatively slow to follow. The Danish Ministry of Research promoted several programmes in the early and mid-1990s, and Australia's Ministry of Science and Technology funded a Proteome Analysis Facility in the mid-1990s. These did much to spark development of proteomic technologies.

But it is only in the past couple of years that funding agencies have generally started to take proteomics seriously as a key post-genomic approach to biological problems.

The UK Biotechnology and Biological Sciences Research Council, for example, has recently funded three centres to establish transcriptomics and proteomics programmes in organisms whose genomes are sequenced, or close to completion. These are the fruitfly *Drosophila*, the yeast *Saccharomyces cerevisiae*, the plant *Arabidopsis* and the bacterium *Streptomyces coelicolor*.

"We see the approach of launching parallel initiatives in transcriptomics and proteomics as one of our most important scientific priorities in the next few years," says Ray

Baker, chief executive of the research council.

In France, five new regional 'genopoles' — sites of concentrated genetics research in a variety of animal, bacterial and plant models — are operating a similar formula. Supported by the research ministry, their funding of FF30 million (US\$4.7 million) each for three years is divided roughly equally between genomics, transcriptomics and proteomics.

The European Commission is also funding a proteomics programme to follow up its yeast genome sequencing programme.

The US Department of Energy recently launched proteomics programmes on lower organisms considered relevant to energy production (methane gas) or the cause of, or cure for, environmental problems (bioremediation). Targeted organisms, again partially or completely sequenced, include several archaea, bacteria-like organisms that often inhabit extreme environments.

The main US proteomics effort, however, is in the health area, with the most significant programmes being coordinated through the National Cancer Institute (see page 718).

Although Britain's Medical Research Council has not created any proteomics programmes, its chief executive George Radda describes the area as "important and exciting", and adds that the council "expects to get lots of applications for proteomics projects through our normal mechanisms".

Learning lessons from NASA's failed science missions

Sir — The probable failure of NASA's Mars Polar Lander, coming so soon after the loss of the Mars Climate Orbiter, will once again call into question the agency's policy of 'faster, better, cheaper' missions (*Nature* **402**, 565; 1999). There is a perception that faster and cheaper is only achieved by taking undue risk. The European Space Agency (ESA) is now changing the nature of its science programme to include faster and cheaper missions, so it is important to understand the pitfalls and benefits of the US experience. Using databases of past US space-science missions, I have assessed the impact of the changed philosophy on the NASA programme.

The number of space-science missions fell systematically from 40 in the eight-year period beginning in 1961 to just 11 in that up to 1992. A projection of the programme evolution made in 1992 would have predicted that, early in the twenty-first century, one space-science mission per decade would become the norm. Clearly, there was a need to reverse the trend: to make missions take place within a shorter schedule and cost less.

Since 1992, the trend has indeed been reversed and there have been 25 space-science mission launches — double the rate in preceding years. But what has been the effect on the failure rate?

As a measure of reliability, I have considered the number of NASA missions for which a failure in the spacecraft brought science operations to a halt earlier than otherwise would have been the case. Such an early end to operations does not necessarily mean that the mission was a failure — often the spacecraft failure came after the prime objectives had been achieved. In the 1960s there was an early end to operations in at least 30% of the missions. Each mission usually involved new technology and at that time there were many lessons to be learned. In the 1970s and early 1980s, an early end to operations occurred in only about 10% of space-science missions. In the years immediately before 1993, the rate was 18%. Since 1992 it will be 24% if Mars Polar Lander is unsuccessful.

A second way of making a comparison is to identify the missions for which it is indisputable that prime objectives were not achieved. In the 1960s, problems occurred before prime objectives were achieved in only 5% of missions. From 1968 to the end of 1992, the failure rate was only 2% (one in 55 missions). For the seven years since 1992, the rate will be 24% if Mars Polar Lander is unsuccessful (six out of 25 missions). If

there is a clear distinction between missions of the 'faster, better, cheaper' era and earlier ones, it is in the increase in the relative number of missions that fail to achieve their prime objectives. Causes of these recent failures include use of a relatively untried launcher, omissions and misinterpretations in the project implementation at several stages, and a failure to convert imperial units to metric.

Before making a final judgement, we should consider the benefits of the new philosophy. The dilemma is that space science carries a high risk, failures are well publicized, and it is less easy to make the same impact with successes. One conclusion is that the US programme has become broader since the change in philosophy. There has been more emphasis on technology demonstration during missions, rather than as a separate activity, and more emphasis on public outreach and education. Broadening the scope of the programme has only been possible because the mission frequency has increased.

For solar terrestrial physics, the emphasis has always been on relatively small spacecraft but in various combinations. Here, the change since 1992 is not dramatic except for a considerable reduction in mission gestation time. For astronomy, the change has been from major missions that one by one extended the wavelength range of observations to smaller missions, each with a more focused set of objectives, building on the observations of the earlier spacecraft. The SWIFT mission is a typical example of a faster, cheaper mission building on discoveries of earlier missions. For planetary science, there has been a major transition from relatively rare but comprehensively instrumented observatories to several much smaller missions with well-focused objectives and with technology demonstration an important ingredient.

To some extent, 'faster, better, cheaper' was a natural next step to follow the major observatory missions of the 1970s and 1980s. But the change in philosophy has not been introduced to the exclusion of the 'slower, expensive' missions: Chandra, SIM, NGST, TPF and Solar Probe all figure in the programme and are the equivalent of missions before 1992. There is no reason to conclude that the changed philosophy has led to a diminution of the science return. On the contrary, while major missions remain part of the programme (though at lower cost), there are now many additional missions, each with short gestation times. Those faster, cheaper missions provide opportunities and science return that was missing from the programme in the 1980s. Perhaps the individual missions may not be 'better' than those of the earlier programmes, but, for the programme overall, 'better' is a suitable description.

ESA has learned from its experience with Cluster to be wary of untried launchers. Fortunately, Europe moved away from mixing metric and imperial units several years ago. The other NASA failures give warnings on how far a 'faster, better, cheaper' approach can be pushed. An advantage that we have in Europe is that there are faster and cheaper programmes running in parallel with the ESA programme. These national missions have a smaller scope than ESA missions and so are implemented more quickly with smaller budgets. These, too, can provide valuable lessons on how far the quest for low cost can be taken. We need to make sure that there is good communication between the ESA and European national programmes, so that we all can learn.

Dave Hall

British National Space Centre, 151 Buckingham Palace Road, London SW1W 9SS, UK

Online research archive will be free to all

Sir — Lance Sultzbaugh criticizes the PubMed Central initiative for having done nothing to change the problems with typical schemes of online scientific publishing (*Nature* **402**, 230; 1999). He says these schemes are difficult to navigate, reside behind cost barriers and conform to no archival standards. It is a little unfair to criticize PubMed Central for being difficult to navigate and lacking in standards before it even exists, but these are certainly valid concerns. It is ridiculous, however, to claim that PubMed Central does nothing to address the cost barriers to accessing scientific literature.

The key characteristic of PubMed Central is that all the research it will archive will be available in full to any individual, at no charge. This applies to research that enters via existing journals that choose to participate, and to research that uses a new peer-review route such as BioMed Central or the Community of Science's scheme.

The PubMed Central model retains a decentralized system of peer review, which is desirable in that no single body can control what gets published. On the other hand, it centralizes the infrastructure necessary to allow trustworthy long-term archiving, effective searching and reliable linking. And it does away with the barriers to access that have prevented the web from achieving its full potential for enhancing scientific communication.

When PubMed Central and BioMed Central are launched, scientists will finally have a viable alternative to traditional journals. Research made available through PubMed Central will be freely accessible,

and yet will also be accepted as a credible, citable contribution to the literature. Surely this deserves to be applauded.

Matthew Cockerill

BioMed Central Ltd, 34–42 Cleveland Street,
London W1P 6LE, UK

How reliable is science information on the web?

Sir — Until recently, the dissemination of scientific information has largely been regulated by publishers via peer review and by librarians through their purchases of journals. With the advent of the World-Wide Web, however, the publication of science has been democratized. Although this ‘deregulation’ will speed the flow of valuable information around the world, a negative side effect may be the increased exposure of students and the public to misleading or biased science, or to opinion masquerading as science. Here, we report on the overall reliability of websites that purvey scientific information.

Our results are sobering. Although the web is increasingly used as a source of scientific information, the quality of the information provided by many of the most easily accessed sites is poor. To gain an understanding of the quality of information on the web, we performed searches for the terms ‘evolution’ (EV), ‘genetically modified organism’ (GMO) and ‘endangered species’ (ES) using Northernlight.com, the search engine with the broadest reported coverage of the web¹.

The first 500 websites retrieved for each topic were examined sequentially by two referees until each had independently reviewed approximately 60 sites containing information pertinent to the topic. These informative sites were scored as ‘inaccurate’ if they contained information that was factually incorrect, ‘misleading’ if they misinterpreted science or blatantly omitted facts supporting an opposing position, and ‘unreferenced’ if they presented information without any peer-reviewed references.

For EV, only 12% (59 of 500) of the websites examined were considered informative by both referees. For GMO and ES, 46% (64 of 140) and 28% (55 of 200) of sites, respectively, were considered informative. Of informative sites, the proportion that were judged inaccurate ranged from 10% for GMO to 34% for EV (Fig. 1). Likewise, the proportion of informative sites scored as misleading ranged from 20% for ES to 35% for EV. A much higher proportion of sites were unreferenced (more than 48% for each category), but the presence or absence of references does not necessarily correspond with the other scores.

Overall agreement values for the refer-

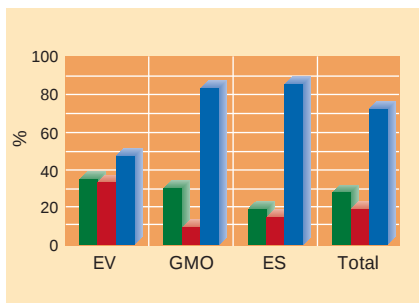


Figure 1 **Percentage of informative websites reviewed for the search topics ‘evolution’ (EV), ‘genetically modified organism’ (GMO) and ‘endangered species’ (ES) that were misleading (green bars), inaccurate (red bars), and unreferenced (blue bars). Data are averaged across referees.**

ees’ scores for the categories of ‘inaccurate’ and ‘misleading’ were 87.8% for EV sites, 82.8% for GMO sites and 73.6% for ES sites. Because the presence or absence of peer-reviewed citations is not subjective, no agreement values were calculated for this category.

Our results indicate that science-related websites have serious liabilities. Many sites purporting to contain science are simply presentations of opinion or social commentary. And the presence of peer-reviewed citations, normally a sign of reliability, does not necessarily reflect the quality of the information presented. Nonetheless, we recognize the substantial advantages conferred by global access to the huge stores of information on the web, particularly for those who might otherwise have limited access to scientific resources.

One promising strategy for such users is the exploitation of recently established portals that provide links to sites that have been reviewed by scientists for accuracy, relevance and currency^{2–4}. These portals, if widely used, also offer a means of establishing peer review as the guiding principle for evaluating science on the web⁵.

Eva S. Allen, John M. Burke, Mark E. Welch, Loren H. Rieseberg

Department of Biology, Indiana University, 1001 E.
3rd Street, Bloomington, Indiana 47405, USA

1. Lawrence, S. & Giles, C. L. *Nature* **400**, 107–109 (1999).
2. Encyclopedia Britannica, Inc. www.britannica.com (4 October 1999).
3. SciCentral. www.scicentral.com/index.html (4 October 1999).
4. Science NetLinks. www.science.netlinks.com/science/index.shtml (4 October 1999).
5. Fainzilber, M. *Nature* **401**, 111 (1999).

Where nuclear weapons come before basic needs

Sir — In his review of the books by Itty Abraham and Michael Foot, Brahma Chellaney chastises Abraham for “his commitment to one side” of the nuclear

debate in India which supposedly “weakens his arguments” (*Nature* **401**, 113–114; 1999). Although Abraham’s sympathies may well lie on the side of India’s growing anti-nuclear movement, Chellaney is very much on the other side.

Chellaney is a member of India’s National Security Advisory Board and an author of the draft “nuclear doctrine”, which calls for “sufficient nuclear weapons to inflict destruction and punishment”¹. So it comes as no surprise that he puts a positive gloss on the history of India’s nuclear weapons, and paints an inaccurate picture of Abraham’s book.

Chellaney persists in describing India’s 1974 nuclear test as a “peaceful explosion”. Such semantic evasions no longer carry any weight, if they ever did. Even Raja Ramanna, former chairman of India’s Atomic Energy Commission and the leader of the team that conducted the test, described it as “a prototype weapon”².

Chellaney also attempts to suggest that India is unique among nuclear weapons states in having straddled the “nuclear fence” for a quarter of a century while having a democratic debate on whether it should “go nuclear”. A less self-serving description would be that India spent decades building a ladder to climb over this imaginary nuclear fence and, when it came time to decide whether to jump down to the other side, a handful of people made the decision, as has always been the case in Indian nuclear policy³.

Abraham’s book shows how Indian scientists used the ideologies of national security and national development to transform a small scientific laboratory into a full-blown weapons complex. His work demonstrates how the combination of science, ideology and the power of the state can be a recipe for disaster as much in the Third World as in the first and second.

Abraham’s real achievement is to reveal the thinking of India’s élite, which places nuclear weapons above providing even the most basic necessities to a large proportion of the country’s citizens.

It is at this level, the right of ordinary people to make meaningful choices about their lives in an informed and democratic way, that Abraham is taking sides. It is a side he shares with Foot, but one far removed from the cabals of “the wizards of Armageddon” who make nuclear-weapons policy around the world.

Zia Mian, M. V. Ramana

Center for Energy and Environmental Studies,
Princeton University, Princeton,
New Jersey 08544, USA

1. National Security Advisory Board *Draft Report on Indian Nuclear Doctrine* http://www.indianembassy.org/policy/CTBT/nuclear_doctrine_aug_17_1999.html
2. Ramanna, R. *Years of Pilgrimage* 100 (Viking, New Delhi, 1991).
3. Ramana, M. V. *Precis* **IX**, No. 3, pp. 1 & 21–25 (1998).

The whens and wherefores of life

A biological slant on the conundrum of our relationship with time.

The Missing Moment: How the Unconscious Shapes Modern Science

by Robert Pollack

Houghton Mifflin: 1999. 240 pp. \$25, £15.99

F. Gonzalez-Crussi

"What is time?" asks Saint Augustine in book XI of the *Confessions*, that obligatory stop, if not point of departure, for every metaphysician bold enough to embark on a voyage in search of the nature of time. And the answer has the same pathetic ring of perplexity today as it did 15 centuries ago, when the saint, who seems to have opened vistas upon every perspective of the problem, wrote: "When no one asks me, I know. But as soon as someone comes to question me on this matter, and I try to explain, I don't know any more..." Indeed, the most treacherous waters in the whole of philosophy's main are those involving time. They who sail forth invariably sink or founder, yet we must salute their gallantry. And this time our excitement is the greater, for biologists have joined the argonauts. Robert Pollack, a professor of biological sciences at Columbia University, is one of them. In his latest book he offers us an inkling of the new ways of thinking that biology will bring to bear on this and other multi-secular quandaries.

Pollack's book may be said to be, among other things, a biologist's meditation on the enigma of time. In the opening chapter, which is devoted to sensation, he tells us that our relational lives are regulated by several 'clocks' that operate at different levels. First, the clock of natural selection, whose ticks, marking the birth of new species, are separated from one another by intervals of millions of years. A second clock, the product of the first, is that of genetic control, whose exquisitely coordinated rhythm in turning genes on or off accounts for normal embryogenesis. A third clock is confined to the nervous system and runs at astonishing speed, its ticks represented by nervous impulses that arrive separated by intervals of the order of thousandths of a second. Our senses, we are told, are assembled by these three clocks. Therefore, the instruments of our cognitive life, however fitted to our purposes they might seem, "were designed to meet the needs of species long dead".

In successive chapters, Pollack deals respectively with the senses, consciousness, memory and the unconscious, microbial infection (a chapter colourfully entitled "The fear of invasion"), cancer ("The fear of insurrection") and death (also "The fear of ...", for symmetry, I suppose). As may be



Time's final enemy: the knight plays chess with Death in the hope of a reprieve from the inevitable.

predicted of a work so bravely taking on the most obstacle-ridden, perplexing quandaries of the human condition, the result is uneven. A clever and engaging expositor, Pollack explains the fundamental time-shift of relational life: there is a lag between the real-time occurrence of a stimulus and our conscious realization or perception of it; we perceive in the present what took place in the past. The brain abolishes the interval, so that perception is experienced as if it were simultaneous with the stimulus.

Outside the realm of the mind, however, the book's focus tends to blur. The author loses none of his narrative flair, and I was delighted to follow his disquisition on vaccines, immunity and microbial life, and the implications of all this for public policy. In his descriptions of infectious diseases, and their multifarious relationships, Pollack lifts the craft of the popularizer to an extended state of perfection. But the reader is unclear on how this ties up with time, and the statement that ends the chapter, that "our sort of internal time does not exist for microbes and never has; they must either divide or die", seems to have been added almost as an afterthought.

Unevenness notwithstanding, *The Missing Moment* is an important book. It invites reflection, and essays of its kind can aspire to no finer encomium. The concluding chapters, which look at the biological processes of ageing and dying, rise beyond mere science vulgarization to a level that may be called,

without hyperbole, true wisdom. The author's treatment of life's final stages — "remarkably like the beginning, the clock of internal time run backward one last time to the earliest days of infancy" — is especially moving. The writer is not afraid of introducing his personal experience, and describes his reaction to these ineluctable events in close members of his family. This directness adds poignancy, and strengthens the plea for scientific research on death and dying that concludes the book.

As a molecular biologist, Pollack often refers to DNA. In addition to the many biological clocks that superintend our existence, he envisages a futuristic "reverse" clock that is rendered possible by thorough knowledge of the genes that increase our susceptibility to disease. Such a clock would measure our age, not in terms of the years we have lived, but of the time we have left. We would be 60 years in the conventional scale, but only five, ten, or one, in the "reverse" measuring system — an interesting idea, to refer time to life expectancy. For time is "a thing that concerns something [else]", as an Aristotelian formula put it. What is this thing? Movement, according to Aristotle; eternity, in the schemes of the ancient philosopher Plotinus; the relationship of the fallen soul to God, said Saint Augustine; the speed of light, observed some modern physicists; DNA, affirm today's biologists. In the last analysis, *The Missing Moment* reminds us, the determinants of time are the

human experiences we all go through: pain, joy, expectancy, desire, disease and, looming behind all these, the inevitability of ever-approaching death. ■

F. Gonzalez-Crussi is in the Department of Pathology, Children's Memorial Hospital, Northwestern University Medical School, 2300 Children's Plaza, Chicago, Illinois 60614, USA.

Numerical concoctions

Once Upon a Number: The Hidden Mathematical Logic of Stories

by John Allen Paulos

Penguin: 1999. 214 pp. £12.99, \$23 (hbk)

Jeremy Gray

Sadly, we do not make our way through life the way professors of statistics, or of any other science, would have us do. Outside the classroom, most of us, most of the time, make elementary errors when reasoning about probabilities. Juries fail to give equal weight to all the evidence and cling obstinately to just those facts that suit their purposes; lawyers have been known to mislead juries over the interpretation of DNA evidence; psychics do well, though doing no better than chance; fraudsters separate the credulous from their money.

To his credit, John Allen Paulos, a mathematician at Temple University in Philadelphia, would not address these problems by herding us all back to school or college. He thinks the problem lies in the way we build up the stories we use to get us through life, the explanations we concoct and the way we feed quantifiable information into the mix. In this book he wanders around the issue, hoping to shed what even he calls an oblique light on the matter.

Sometimes the issue is straightforward; seemingly very rare occurrences, in fact, crop up all the time. We have a bias towards blaming people when someone is injured, but saying it was only bad luck when an exactly similar mistake injures no one. In such cases, our prior views about the way the world operates twist our use of the numbers.

Everyone uses rules of thumb to deal with the complexities of real life. These may have something going for them, even if they wouldn't get good marks on a statistics course, because they allow us to filter out the complexity. Here, however, the author may not appreciate the scale of the problem. At one point Paulos jokes (and the book is full of jokes, some of them very funny) that the typical statistics problem vexes thousands and pleases seven or eight. In fact, most of his examples are contrived in this possibly vexatious way. This is partly

because Paulos is writing a popular book in which the arithmetic must be made easy and the assumptions kept simple. But it is partly a function of the way statisticians operate, because the complexities of real life are daunting. The end result, nonetheless, is hypothetical situations that are too simple to convince us that their statistical morals apply to real life, and it is not always clear that Paulos realizes this.

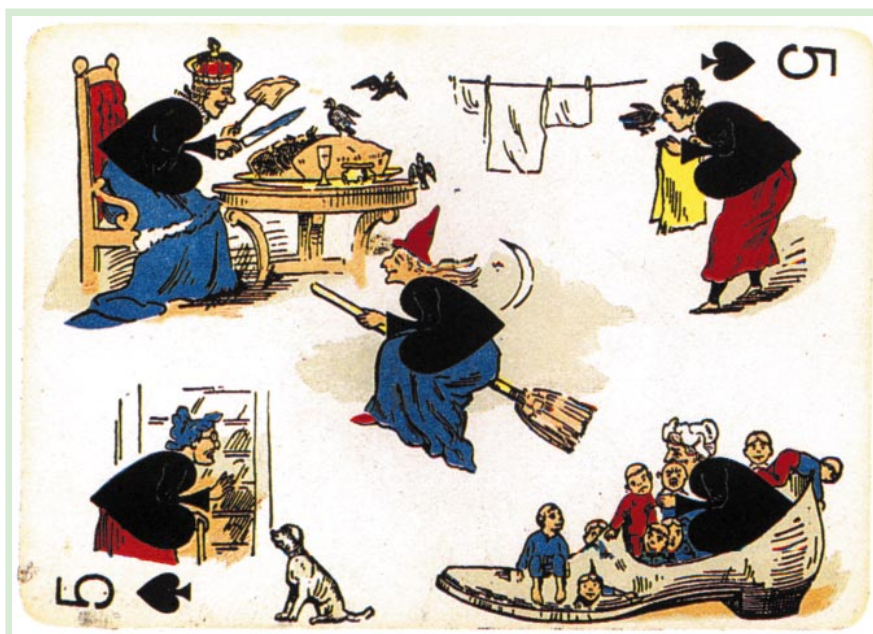
Statistics is not the only discipline to take a toy fishing-net to the ocean. The second half of the book is at least as interested in the disparity between simplified pure logic and the ways we reason, which various novel accounts of logic attempt to describe. Paulos mentions situational logic in this context. He also considers the complexity of the world, describing how complexity theorists think about it; he alludes to Ramsey theory, and generally discusses how order and simplicity always seem to turn up somewhere.

These are important issues, and the reader who does not know of this work will be amused and diverted. But the claim that this connects to our use of stories is a tenuous one. It is not that anything Paulos says is wrong, so much as that everything he says is superficial. There are deep problems here about order, chaos and complexity, on the one hand, and how our mind works, on the

other, but on the evidence presented here we are very far from putting the two together.

Paulos intends the term 'story' to mean more than just the little analyses we make up for ourselves. He wants it to reach as far as published stories, from romantic and detective fiction to the profundities of Anton Chekhov. There is nothing here that will help the reader enjoy fiction more, read it more carefully or understand how it works. We are indeed dangerously close to the literal-minded approach to fiction so ruthlessly parodied under the heading, "How many children had Lady Macbeth?"

The book comes nowhere close to living up to its subtitle. It does not provide a mathematical logic for stories, but merely suggests that intentional logic addresses the issue of finding meanings in the confusing world around us. It does convey a good sense of the difficult nature of that task, at which humans manage to be remarkably good while being in some basic ways rather bad. There is probably somewhere an automatic theorem-proving machine that can outstrip most of us at solving puzzles of a logical or probabilistic kind. But if so, that would only prove yet again that computers can do some things very much better than we can, and still can't think. Resolving this conundrum will help us understand how the mind works. Until then,



Playing with numbers

It may look like a nursery rhyme montage, but the illustration above is actually the five of spades from a surreal pack of playing cards. The challenge in producing such 'transformation' cards lies in intriguing the eye with some kind of visual pun or surprise while minimizing the element of distraction from the card's suit and value. The first such playing cards were produced in Germany in 1801, and pictorial

themes have since ranged from the satiric and comic to the erotic and commemorative. Illustrations of other transformation cards form just one feature among the many tantalizing images to be found in *The Playful Eye: An Album of Visual Delight* edited by Julian Rothenstein and Mel Gooding (Redstone, £19.95; pbk), a collection of popular graphics of the nineteenth and early twentieth centuries.

books like this one will tell us, very enjoyably, that there is work to be done. ■

Jeremy Gray is in the Department of Mathematics, Open University, Milton Keynes MK7 6AA, UK.

More on numbers

What is Random? Chance and Order in Mathematics and Life

by Edward Beltrami
Springer, \$22, £15.50

Imaginary Numbers: An Anthology of Marvelous Mathematical Stories, Diversions, Poems, and Musings

edited by William Frucht
Wiley, \$27.95, £22.50

Light from underground

Mosaic Evolution of Subterranean Mammals: Regression, Progression and Global Convergence

by Eviatar Nevo
Oxford University Press: 1999. 413 pp.
£95, \$175

Hynek Burda

Across the globe, at least 285 of 4,629 species of mammals, representing 11 families, spend most of their lives in moist, dark, oxygen-poor and carbon-dioxide-rich burrows, deprived of most of the sensory cues available above ground. These mammals have become fully specialized for a unique way of life in which foraging, mating and breeding take place underground.

Although most subterranean species have been known to scientists for a long time, their biology has remained largely unstudied. This may be explained by their cryptic way of life and the related problems of keeping, breeding and monitoring them, and also by the fact that scientists tend to be more attracted by animals confronting environments and problems above ground that seem far more complex than those encountered below ground (sensitive vision compared with blindness, echolocation compared with human-like hearing, long-distance navigation/maze orientation across tens of metres, thermoregulation in the cold/life in a thermally buffered burrow). Although many specimens of moles (insectivorous subterranean mammals) and mole-rats (subterranean rodents) have been deposited in museums, not even the morphology of their digging specializations has received the attention it deserves. The convergent evolution of subterranean mammals, one of the most remarkable examples of convergence, is rarely mentioned in textbooks.

The 'sleeping beauty' of subterranean

New in paperback

Constructing Quarks: A Sociological History of Particle Physics

by Andrew Pickering
University of Chicago Press, \$26, £18.50

Great Feuds in Science: Ten of the Liveliest Disputes Ever

by Hal Hellman
Wiley, \$15.95, £9.99

Science As A Way of Knowing: The Foundations of Modern Biology

by John A. Moore
Harvard University Press, \$18.95, £11.95

Warmth Disperses and Time Passes: The History of Heat

by Hans Christian von Baeyer
Random House, \$13.95, £11.99

The Gospel of Germs: Men, Women and the Microbe in Modern Life

by Nancy Tomes
Harvard University Press, \$16.95, £10.50

Making Sense of Illness: Science, Society and Disease

by Robert A. Aronowitz
Cambridge University Press, £11.95, \$29.95

Nature Wars: People Vs. Pests

by Mark L. Winston
Harvard University Press, \$15.95, £9.95

The Handicap Principle: A Missing Piece of Darwin's Puzzle

by Amotz & Avishag Zahavi
Oxford University Press, £11.99, \$16.95

Consilience: The Unity of Knowledge

by Edward O. Wilson
Abacus, £8.99, \$14

The Woman That Never Evolved

by Sarah Blaffer Hrdy
Harvard University Press, \$16.95, £10.50

Social Mindscales: An Invitation to Cognitive Sociology

by Eviatar Zerubavel
Harvard University Press, \$15.95, £9.95

Blind Watchers of the Sky

by Rocky Kolb
Oxford University Press, £8.99

The Anti-Depressant Era

by David Healy
Harvard University Press, \$17.95, £10.95

mammals, and their importance for science, has been awoken by Eviatar Nevo of the University of Haifa, Israel. His contribution to the growth of our knowledge of subterranean mammals has not only inspired others; he has himself co-authored at least 20 per cent of all published studies.

Growing knowledge, and the landmark dates of Nevo's major contributions (1969, 1979, 1989), seemed to predestine 1999 as a further important year, and the time was right for his book to appear. Undoubtedly, Nevo is the most competent person to write it. And the expected monograph, which went unpublished for several decades, has at last appeared.

The book describes and analyses the 40 million years of global evolution of subterranean mammals and its implications throughout biology. Although the underground habitat is in many respects relatively simple, monotonous, stable and predictable, it is in others very specialized and stressful. Consequently, the evolution of subterranean mammals involves dramatic and complex adaptive structural and functional changes that are both regressive (degenerative) and progressive (compensatory). This mosaic convergent global evolution of subterranean mammals is an example *par excellence* of comparative studies in evolution at all organizational levels, from the molecular to the

organismal, oriented by natural selection.

The book is filled with information. The reader will find not only an up-to-date overview of subterranean mammals and their evolutionary problems, but also detailed information and references to general aspects of sensory and behavioural ecology, morphology, physiology, genetics and immunogenetics. This is all based on Nevo's 50 years of studies and experience, and on careful, critical and thoughtful study of hundreds of articles and books. The book provides excellent texts for seminars and courses. It is richly illustrated, and there are more than 1,800 entries in both the index and the well-balanced bibliography. The book is a 'must' for all students of subterranean mammals, and will be very useful to evolutionary biologists.

Considering the hitherto catalytic effects of Nevo's work, I would like to bet that the number of studies dealing with subterranean mammals will rise dramatically in the coming years (indeed, there is great potential for surprising discoveries), and that there will quickly be a need for a second edition. No doubt, subterranean mammals will soon be burrowing their way into the textbooks of the future. ■

Hynek Burda is in the Department of General Zoology, Faculty of Biosciences (FB 9), University of Essen, D-45117 Essen, Germany.

Moving from simple to complex

Neuronal Control of Locomotion: From Mollusc to Man

by G. N. Orlovsky, T. G. Deliagina and S. Grillner

Oxford University Press: 1999. 322 pp. £49.50

Nicholas Dale

The triumvirate of Orlovsky, Deliagina and Grillner is uniquely placed to produce a unified account of the neural basis of locomotion across the whole animal kingdom. Their collective experience in the field of locomotor control spans more than 30 years. All have made major contributions to understanding the control of movement in the cat, before moving to simpler preparations in search of more detailed understanding.

In surveying this vast field, the authors systematically describe the mechanisms underlying locomotion in a range of invertebrate and lower vertebrate preparations, moving to a detailed account of mammalian locomotion. The book's great strength is that, for each preparation, it starts with the basics and leads the reader to a fairly good understanding at the levels of systems and networks. By drawing parallels and contrasts between work done in different preparations and contexts, the authors avoid compartmentalizing the field. For example, they highlight the parallels between the mammalian and arthropod limb controllers, and describe clearly the problems of effective locomotion and the strategies used by the nervous system to overcome them.

There is a lack of simple diagrams that would have made some ideas more accessible. Nevertheless, this is an excellent book for those seeking wide background knowledge of neural mechanisms underlying movement. Those already in the field will also find it valuable. At times, however, there is a lack of balance. Perhaps understandably, the authors' own work receives detailed treatment. But this comes at the price of incomplete coverage of more recent developments in neuromodulation and the contribution of ion channels to network function. The chapters on mammalian locomotion concentrate almost exclusively on the cat, with only a passing mention of findings in the rat. The mouse and rat will surely be important models in the post-genomic era, and deserve fuller treatment.

Nevertheless, while it is easy to criticize the difficult choices inherent in writing this book, the authors have succeeded in illustrating the progress that has been made by studying simple systems, and how this illuminates the principles underlying locomotion in more complex animals, including man. ■

Nicholas Dale is in the School of Biological and Medical Science, University of St Andrews, Fife KY16 9TS, UK.

Science in culture



Photography and the artist

"Degas to Picasso: Painters, Sculptors and the Camera", a travelling exhibition

Marina Vaizey

This major exhibition juxtaposes photographs as they were actually taken, or found and used, with original paintings, sculptures, drawings and prints by 14 European artists from the 1860s to the first decade of the twentieth century. The purpose is to demonstrate the complex relationships between the mechanically taken image and the image created manually by the same artist. The relationship may be direct, as when the Norwegian expressionist painter Edvard Munch based painted compositions directly on photographs, or oblique, as when the *intimiste* painter Edouard Vuillard, enthralled by his Kodak, took countless photographs of his friends, family and milieu, which informed his paintings without being directly used (see above).

Curiously, the American, Thomas Eakins, who was so resonantly involved with photography, not only taking his own photographs but also working directly with Eadweard Muybridge, the pioneer of action photography, is omitted. This is evidently because a thesis of the compilation is that photography was as important for symbolist aesthetics as for nineteenth-century realism. But the importance of photography for a hugely stylistically and philosophically diverse group of artists has been known since the relationship between the photographer and the painter was first studied seriously some 40 years ago. And such an omission throws up another problem: the quality of the artists on display is so uneven that the compilation at times lacks gravitas.

What, however, does this exhibition show us, and the expansive accompanying publication of the same title (Yale University Press, \$50, £35) tell us, about the artist and the camera? The main point for all those who are sceptical about photography is how much more it is than a mechanical tool. Light is shed on the work of each artist merely by absorbing what photographs they chose as inspiration, *aide-mémoire*, or sketch and what kind of photographs they themselves took. The two best-known artists under examination, Edgar



Edouard Vuillard used his photograph of Misia Natanson (left) indirectly in his painting *Mother and Child*.

Degas and Pablo Picasso, had a complex relationship with photography which has only recently come under sustained scholarly scrutiny. The exhibition gives a succinct presentation for Degas. His favourite subjects included horse-racing and ballet dancers, and Muybridge's photographs of horses in motion and dancers in movement, from his famous publication *Animal Locomotion*, are included next to Degas' own sculpture of horses and paintings, sculptures and drawings of dancers.

The representation of each artist and related photographs is necessarily thin, and the reading matter required, in terms of effective exhibition design and display, excessive. Not for the first time do we find a far better and more interesting publication than exhibition.

In 1839, the French artist Paul Delaroche, perhaps the most influential depicter of historical scenes in the Europe of his period, rather dramatically announced that, with the invention of photography, painting was dead. Far from it: the camera simply became one more weapon in the armoury of the painter and sculptor. Not only have artists quarried the riches offered by the photographic image, but the techniques of photography — for example, the cut-off figure so characteristic of the snapshot — have been reinterpreted in paint. Moreover, artists rapidly adopted photography as a new kind of sketchbook. It has been argued that photography was also responsible, because of its apparent ability to portray the external world objectively, for allowing the horizons of painting and sculpture to expand into abstraction. This seemingly new vocabulary of forms has been called the non-objective world, that is, the world of the imagination rather than that perceived by the senses. ■

Marina Vaizey is a writer and art critic.

The exhibition is at the San Francisco Museum of Modern Art, San Francisco, California, until 2 Jan. 2000; at the Dallas Museum of Art, Dallas, Texas, 1 Feb. – 7 May; and at the Guggenheim Museum, Bilbao, Spain, 12 June – 10 Sept.

Slanted evidence

The tortuous path from skull measurements to theories of racial superiority.

Martin Kemp

We have inherited a notably persistent image of 'primitive man'. 'He' is characteristically hairy, raw boned, mighty in hand and foot and endowed with a prognathous jaw. The basics of this stereotype were laid down with the 'wild' or 'savage' races believed by the ancients to inhabit regions remote in place or time. The new factor in the nineteenth century was the simian nature of his cranium and face. Most specifically, primitive character was signalled by a facial angle that ran obliquely backwards from protruding jaw to receding forehead.

Such an ape-like physiognomy can most obviously be recognized as an outcome of Darwin's theories of human evolution. However, facial angle had been established as a measure of beauty towards the end of the preceding century, above all in the widely read publications of Petrus Camper, the learned Amsterdam doctor. In his *Dissertation sur les Variétés Naturelles qui Caractérisent la Physionomie des Hommes*, published in 1791, and translated in *The Works of the Late Professor Camper, on the connection between the science of anatomy and the arts of drawing, painting, statuary &c* in 1794, he noted that "it is amusing to contemplate an arrangement ... in a regular succession: apes, oranges, negroes, the skull of a Hottentot, Madagascar, Celebese, Chinese, Moguller, Calmuk and diverse Europeans". This was precisely how he had arranged his own collection of crania "on a shelf in my cabinet".

Using an elaborate measuring frame of his own devising, Camper tabulated and illustrated progression of facial angles: the skull of a tailed monkey measured 42°; that of a small orang-outang (on which Camper had written a monograph) was 58°; a young negro exhibited a 70° inclination; a Calmuk ("deemed the ugliest of all the inhabitants of the earth") weighed in at the same angle, whereas the typical European profile attained the near-vertical with 80°. This peak could be surpassed "by the rules of art alone", as in the lauded ancient sculpture of the Apollo Belvedere which stands at the head of Camper's sequence.

In retrospect, it is all too easy to see how this succession could be used to align apes and human races in a moral order which placed the 'negroes' at the baser animal end of the scale. In conjunction with the voguish physiognomics of Johann Kaspar Lavater, this diagnostic move was readily made, not least in Charles White's *An Account of the Regular Gradations in Man and in Different*

Kinds of Animals and Vegetables of 1799, in which profile drawings of heads were arranged in ascending succession from a long-beaked bird (literally bird-brained) to the noble profile of an exemplary Caucasian. However, Camper himself was not prepared to make this move.

Camper rejected the "extravagant" notion that "that the race of blacks originated from the commerce of the whites with oranges and pongos; or that these monsters, by gradual improvements, finally become men". The morphological differences between quadruped apes and upright men "seem to mark the boundaries which the creator has placed between the various animals". The "divergences" persuaded the anatomist that "the whole human race as it is now spread over the face of the earth" was originally "descended from a single pair, that were formed by the immediate hand of God, long after the world itself had been created and had passed through numberless changes". While Special Creation or a sequence of Spe-

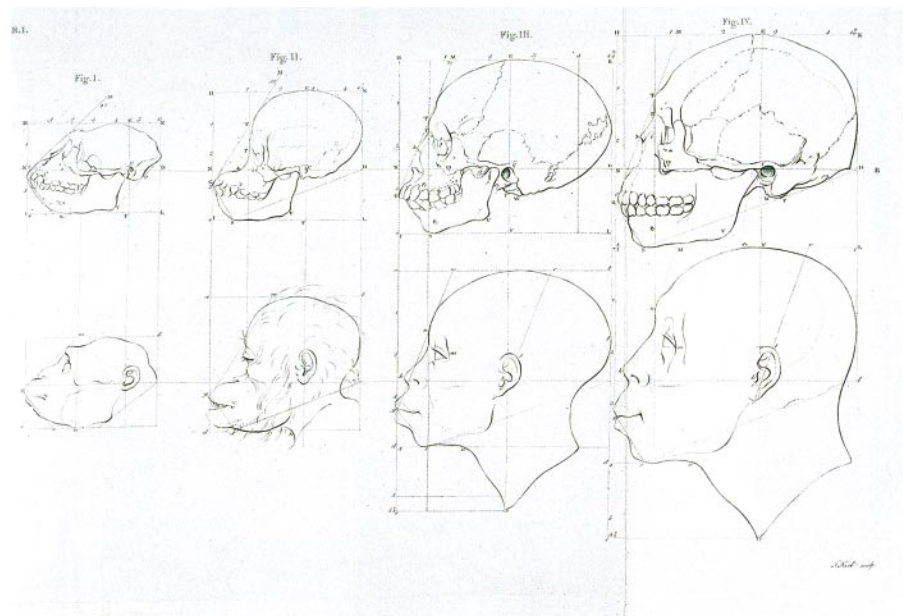
cial Creations remained the norm, there was a limit to the conclusions that could be drawn from any supposed similarities between the skulls of 'negroes' and apes.

One consequence of the Darwinian revolution for craniometry was to arrange the Camperian sequence of angles in a temporal succession. Once the steps represented by God's discrete creation of separate species had been replaced by the transformation of one species into another — or monkeys into men, as was popularly said — the kind of fallacious ranking effected by White was granted a new scientific licence. Camper's 'negro' could be stigmatized as evolutionarily more primitive. The nineteenth-century campaigns of craniological measurement and ethnographic photography, together with huge collections of skull types, provided great banks of data that could be exploited to detect those races that stood closest to the animal origins of man.

The progressive misuse of Camper's non-racist characterizations is a salutary reminder of what can happen to apparently 'neutral' results when a climate of belief is radically transformed. Place his observations together with Lavater's diagnostic physiognomics and Darwin's *Descent of Man* in the same social cooking-pot, and we can see how an ugly dish can result. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK.

Heads were arranged in ascending succession from a long-beaked bird to a noble Caucasian.



Class of the head: Petrus Camper's "amusing" succession of skulls from apes to man.

The road to the year 3000

The simplicity of a single equation holds the key to human happiness.

Harry Harrison

It has certainly been an interesting third millennium. High points that spring to mind are the intergalactic wormhole expeditions — all 13 of them. All departed as planned. Of course, none of them has returned yet.

Perhaps more satisfying has been the global reduction of greenhouse gases. So successful has this been that, in fact, the global ice caps are growing and the glaciers expanding. The new industry of growing forests, just to burn them for the carbon dioxide that they release, is becoming a most profitable one worldwide.

Most interesting, perhaps, was the discovery in 2688 of the remains of germanium-based life-forms that might possibly have visited the planet Pluto. This is still very much of a mystery. Was it hoax or visitation? The jury is still out.

But these, and other physical events of the millennium, are far outweighed by the idea, the theorem, the equation — the concept — that has changed every element of our society, every formula, every scientific discovery that makes mankind what it is today.

At the risk of being pedantic, I draw your attention to the Universe. It exists. It functions. Interactions occur at every micro and macro level. Scientists observe, study — and discover. The animals of the Galapagos Islands had millennia to mutate before Charles Darwin arrived. It was his intelligent observations of them, then his ratiocination, that produced the *Origin of Species*.

Albert Einstein did not, of course, invent energy, for it existed independently of him, and was there for him to study — waiting, some might say, for his application, clarification and classification. He possessed the skill, the intuition, the intelligence to observe and simplify — and declare that $E = mc^2$.

These are two samples out of thousands — hundreds of thousands — that clearly demonstrate that it is the application of intelligence to observation that reveals nature's secrets. But, oh, it is such a haphazard occurrence that we must stand in awe at how much has been learned in such an unstructured manner.

Now, as we approach the end of this third millennium, and look forward to the as-yet-unrevealed wonders of the fourth, we must bow our heads in gratitude to the man and the woman who discovered, and formula-ized for mankind, exactly how the process of discovery works.



We are all familiar with the titanium sculpture of this couple, standing in the station of the lunar shuttle in Mare Serenitatis, where they met. Then, because of a photon storm, the shuttle was one hour late, and they talked.

Stern was a professor of philosophy, specializing in intuitive logic; Magnusson a physicist well known for his study of tachyons. They seemed to have little in common, other than an academic background. Nothing could be further from the truth; their respective disciplines embraced each other like the Yang and Yin.

We will never know those first words that they spoke to one another. Would that we could! We must settle for the scribbled equations on the back of an envelope: equations that poured from the fruitful mating of those two great minds. Before their shuttle trip was over, the basic theory was clear, the applications virtually universal. Before the next day was out, the equations were clarified, reduced, and finalized into the Stern-Magnusson Equation as we know it today.

Even this genius couple stood in awe of what they had created. Magnusson applied the formula to a problem of tachyon spin

that he had been worrying at for more than a year. It was solved on the spot.

Almost as an attempt at humour to alleviate the intense emotion of the occasion, Stern wrote out the equation and filled in the elements for the never-solved relationship between sunspot cycles and the birth-weight of male Greenlanders. And gasped. There was the answer — clear and obvious now that the equation had been solved.

The rest is history. With these seven mathematical symbols, mankind has regularized the dictionaries of the Anglo-Saxon language, predicted earthquakes and tsunamis, found hidden oil reserves, abolished city traffic jams — and moved on from the glorious past to the even more glorious future. No scientific discipline has resisted the irresistible logic of this equation. The Stern-Magnusson Equation is the definitive discovery of humanity, the finding that has freed mankind.

Of course we have all learned it in school. But it always bears repeating, for we can never tire of it. The Stern-Magnusson Equation goes like this...

Harry Harrison has published science fiction for 50 years. He is the author of 44 novels, the latest of which is *The Stars and Stripes Forever*.

JACEY

Today we have naming of parts

Elliot M. Meyerowitz

The first completely sequenced plant chromosomes, from the mustard *Arabidopsis thaliana*, reveal a dynamic genome that is constantly being rearranged.

Although we share the planet with plants and depend on them for food, raw materials and energy, we know very little about them. But this is beginning to change. On pages 761 and 769 of this issue, Lin *et al.*¹ and Wambutt *et al.*² report the DNA sequences of chromosomes 2 and 4, respectively, from *Arabidopsis thaliana*. The sequences include about 30% of the plant's genes, and show that although many plant genes are familiar homologues of known animal and fungal genes, many others are plant specific. The sequences also show an evolving genome, with evidence for ancient and recent duplications of both single genes and large chromosomal regions.

Arabidopsis (Fig. 1) was chosen almost a decade ago as the subject of the first plant genome project because it has a very small genome, yet performs the same functions and contains essentially the same genes as other flowering plants. In other words, it has the same list of parts as more familiar crop and ornamental plants, yet lacks the large arrays of repetitive DNA sequences found between genes that characterize larger genomes. *Arabidopsis* is the main model system for laboratory studies in basic plant biology, so we know much about its genetics, physiology, development and structure³. Its nuclear genome contains five chromosome pairs, and Lin *et al.* and Wambutt *et al.* have now sequenced the two smallest, chromosomes 2 and 4. Their sequences are complete, with the exception of the region around the centromeres (a specialized structure important in cell division), which does not contain many genes. The total of 37 megabases (Mb) is more than a quarter of the *Arabidopsis* genome, and the rest is expected to be done within a year.

What is found in this list of parts? The authors have identified 7,781 protein-coding genes. Because the gene-rich part of the whole *Arabidopsis* nuclear genome is around 120 Mb, 31% of genes ought to be in chromosomes 2 and 4, meaning that the entire genome should code for around 25,000 proteins. This is a surprisingly large number compared with the predicted totals of fewer than 20,000 genes in the nematode worm *Caenorhabditis elegans*⁴ and fruit fly *Drosophila melanogaster*⁵ genomes. With their nervous systems, behavioural responses and ability to move (plus, in one case, to see and fly too), they seem much more



Figure 1 The tiny mustard *Arabidopsis thaliana*. (Painting by Janet Wehr, courtesy of Chris Somerville.)

complex than *Arabidopsis*, which has only around 50 identified cell types. But this apparent greater complexity may be an illusion, as plants do many things that animals cannot. As well as producing oxygen by photosynthesis and synthesizing all their own amino acids and vitamins, plants can assess (and respond to) environmental features such as gravity, day length and the quality of light.

The evolutionary lineages leading to plants and animals diverged around 1.5 billion years ago, when the last common ancestor was some unknown sort of unicellular organism⁶. All the functions of a multicellular, land-dwelling organism — such as cell–cell communication and water uptake — then evolved independently in the two lineages. This is why plants and animals are so different anatomically. But how different are they at the level of their gene lists?

Gene function can be predicted, to a certain degree, by comparison with known gene sequences. For almost 60% of the *Arabidopsis* proteins encoded on chromosomes 2 and 4, the functions (including plant- and bacteria-specific processes such as photo-

synthesis, as well as more general processes shared with animals) are familiar. For example, 5% (218) of the 4,037 *Arabidopsis* genes on chromosome 2 code for DNA-binding proteins, which are implicated in gene expression in both plants and animals. Some of the families of such proteins found in *Arabidopsis*, such as myb relatives and homeobox proteins, are also present in animals. Other families, though, are so far found only in plants⁷.

Similarly, animal and plant genomes code for thousands of receptor proteins. These proteins, which can be at the cell surface or internal to the cell, allow cells to sense signalling molecules such as hormones. In plants, like animals, such cell-surface receptors are often protein kinases. But those found so far in plants have all been serine/threonine or histidine kinases, whereas tyrosine kinases are usually found in animals. So although the general function of many genes seems to be similar in plants and animals, in many instances the related function is carried out by a protein with a different evolutionary history.

Why is it important to know this? First, because comparisons between plants and animals can show which general functions are essential for complex organisms, and which different proteins can carry out these functions. And second, because the availability of plant proteins and plant pathways that can be transferred to transgenic animals (where they could act without interacting with existing animal pathways), and vice versa, gives a large palette of genes for experimental use. If almost 60% of the genes can be assigned a general function by comparison with known proteins, this means that more than 40% cannot. Many of these are specific to plants, so there are also hitherto unknown protein functions — and, perhaps, unknown functions of organisms — waiting to be discovered.

As well as the list of parts, the chromosomal sequences reported by Lin *et al.*¹ and Wambutt *et al.*² show some of the processes that lead to large-scale genome structure. Both chromosomes show many tandem gene duplications (where two or more close copies of a gene are found next to each other). For example, on chromosome 2 there are 239 tandem duplications, involving 593 genes. There are larger and more distant duplications too — almost 2.5 Mb of

sequence (with considerable divergence) is duplicated on both chromosomes in four large blocks. Moreover, a portion of chromosome 4, containing 37 genes, is duplicated on chromosome 5.

Chromosome 2 also contains a region equivalent to 75% of the mitochondrial genome. (Mitochondria, where respiration takes place, have their own genome, which is thought to have arisen through engulfment of a bacterium.) This region of chromosome 2 is nearly identical to the mitochondrial genome, indicating the recent transfer of a large piece of DNA from a membrane-bound organelle to the nucleus. Overall, more than 60% of the predicted proteins have a relative with considerable sequence conservation elsewhere in the *Arabidopsis* genome. The degree of sequence conservation between duplicated genes and regions indicates that large and small regions have duplicated, both locally and distantly, over the course of chromosome evolution in this tiny mustard.

This first glimpse of plant chromosomes — and the complete lists of the genes on them — shows a remarkably dynamic genome, one in constant motion and under-

going constant rearrangement. Nonetheless, there is a large enough core set of genes similar to those of distant relatives (such as animals) that functional conservation seems likely. Moreover, there are enough related functions to see that the independent solutions to the problems of multicellular existence found in plant and animal lineages can involve similar biochemistry. Such reactions are carried out sometimes by related, and other times by unrelated, proteins. Finally, there is a large collection of genes that do not correspond to anything yet seen — enough to guarantee many years of discovery in plant science. ■

Elliot M. Meyerowitz is in the Division of Biology, California Institute of Technology, Pasadena, California 91125, USA.
e-mail: meyerow@caltech.edu

1. Lin, X. *et al.* *Nature* **402**, 761–768 (1999).
2. Wambutt, R. *et al.* *Nature* **402**, 769–777 (1999).
3. Meyerowitz, E. M. & Somerville, C. R. (eds) *Arabidopsis* (Cold Spring Harbor Laboratory Press, New York, 1994).
4. *C. elegans* Sequencing Consortium *Science* **282**, 2012–2018 (1998).
5. Miklos, G. L. G. & Rubin, G. M. *Cell* **86**, 521–529 (1996).
6. Wang, D. Y. C., Kumar, S. & Hedges, S. B. *Proc. R. Soc. Lond. B* **266**, 63–171 (1999).
7. Riechmann, J.-L. & Meyerowitz, E. M. *Biol. Chem.* **379**, 633–646 (1998).

Extrasolar planets

Direct detection at last

Adam Burrows and Roger Angel

In the past four years, astronomers have discovered at least 30 planets in orbit around nearby stars^{1–11}. Their detection relies on the fact that the force of gravity that keeps a planet in its orbit has an opposite reaction on the star, causing it to wobble in its own tiny orbit about the common centre of mass. From the size of the stellar wobble, astronomers can calculate a lower limit for the planet's mass, but they can't tell whether they are seeing the effect of a lighter planet in an orbit seen edge-on, or a heavier one in the same orbit viewed differently — until now. On page 751 of this issue, Cameron *et al.*¹² argue that they have detected light reflected from a planet orbiting a nearby star, providing a direct measure of its mass and size.

All the stellar wobbles indicating extrasolar planets have so far been found using the Doppler technique, which measures changes in the star's velocity along the line of sight. What is actually measured is a small, systematic change in wavelength (Doppler shift) of the many absorption lines that make up the star's spectrum. Jupiter, for example, causes the Sun's velocity to vary by $\sim 12.5 \text{ m s}^{-1}$ during its 12-year orbit, whereas a planet much closer to its parent star produces a larger velocity perturbation (in the case of the planet orbiting 51 Pegasi, $\pm 60 \text{ m s}^{-1}$). The

Doppler method favours systems quite unlike our own, with heavy giant planets close to their stars producing the strongest and most easily detected gravitational tugs. Some have orbits as short as a few days, rather than decades, and many extrasolar planets are heavier than Jupiter. Although these discoveries have serious implications for theories of planet formation, the indirect detections cannot tell us anything about the radius or composition of these planets.

The planet detected by Cameron *et al.*¹² is orbiting every 3.3 days around the star τ Boötis, which is 50 light years away near the constellation Boötes. The planet's tight orbit causes a relatively large stellar Doppler shift of $\pm 470 \text{ m s}^{-1}$, implying a mass 3.9 times or more that of Jupiter. The planet is so close to τ Boötis (20 times closer than the Earth is to the Sun) that even the Hubble Space Telescope can't distinguish them. The combination of close orbit, high mass and the fact that τ Boötis is hotter and bigger than our Sun, means this planet is one of the hottest known.

The planet must be brilliantly illuminated, and is a prime candidate for direct optical detection. Although it will reflect to us only about 0.01% of the total starlight from τ Boötis, this light will be strongly Doppler

shifted because of the planet's high speed. Cameron *et al.* find evidence for exactly such an effect, a faint planetary spectral component that is strongly blueshifted by $74,000 \text{ m s}^{-1}$ when the star is redshifted by 470 m s^{-1} (redshift indicates an object moving away from us, whereas blueshift indicates an object approaching Earth). This result directly reveals that the planet weighs $470/74,000 \times (1.25 \text{ solar masses}) = 8$ Jupiter masses (M_J). This mass is consistent with the lower mass limit from the amplitude of the stellar wobble, provided that the plane of orbital motion is being viewed from above, with the axis of rotation tilted 29° in our direction.

The strength of the planet's signal tells us about its size and brightness. Assuming an albedo (the amount of light reflected) of 55%, similar to that of Jupiter, Cameron *et al.* deduce that the planet must be 1.6–1.8 times bigger than Jupiter — that is, 30–50% larger than expected from theory¹³. A smaller albedo would imply an even larger radius. In addition, the authors find the planet to be blue–green in colour, something mildly at odds with emerging theories of the reflectivity of extrasolar planets^{14–17}. So, if true, the Cameron *et al.* findings will demand new explanations.

Unfortunately, the detection of reflected light is extremely difficult and so reliability is not high. Most worryingly, when the data are analysed at orbital periods other than the known period of the planet, the authors find spurious signals similar to (or larger than) that taken to be the planetary signal. Indeed, there is a 1 in 20 chance that the apparent planetary signal is just random noise. It is also true that previous independent observations¹⁸ did not find a planetary signal from the same star and in the same region of the spectrum where Cameron *et al.* find their strongest result. We expect that more extensive observations made next spring, when τ Boötis is once again high in the night sky, will resolve the issue.

In the meantime, unimpeachable direct evidence of an extrasolar planet has been found in the light from star HD209458 because of the transit of a planet across its stellar disk^{19,20}. At the moment when the star is furthest from us, the starlight is seen to dim by about 2% for three hours, as the planet passes in front. To see a transit we must be viewing this system nearly edge-on, so the planet's mass must be the minimum given by the star's wobble, $0.63 M_J$. From the amount of dimming, and the known size of the star, the planet is found to be 1.27 times the radius of Jupiter (R_J), independent of any assumptions about albedo. This is in harmony with theoretical predictions¹³ of between 1.2 and $1.6 R_J$ for a fast-orbiting Jupiter-like planet. Any last doubts about the existence of Jupiter-mass planets are removed by this direct measure of mass and size. Not only is

the planet real, we now know that it is a gas giant made mainly of hydrogen, not rock—a solid planet with a mass of $0.63 M_J$ would have to be three times smaller than Jupiter.

Our understanding of planetary science will be enormously increased by direct measurements of the physical properties of extrasolar planets. The detection of reflected light from a planet, if confirmed, is the first step towards spectroscopic studies of extrasolar planets that would enable us to study their chemical composition and atmospheres. A journey that begins with giant planets close to their primary stars should in time lead to a better understanding of Earth's place among the many planetary systems that seem to surround us. ■

Adam Burrows and Roger Angel are in the Department of Astronomy and Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA.

e-mails: aburrows@as.arizona.edu
rangel@as.arizona.edu

1. Butler, R. P. & Marcy, G. W. *Astrophys. J.* **464**, L153–L156 (1996).
2. Butler, R. P., Marcy, G. W., Williams, E., Hauser, H. & Shirts, P. *Astrophys. J.* **474**, L115–L118 (1997).

3. Butler, R. P., Marcy, G. W., Vogt, S. S. & Apps, K. *Publ. Astron. Soc. Pacif.* **110**, 1389–1393 (1998).
4. Cochran, W. D., Hatzes, A. P., Butler, R. P. & Marcy, G. W. *Astrophys. J.* **483**, 457–463 (1997).
5. Fischer, D., Marcy, G. W., Butler, P., Vogt, S. S. & Apps, K. *Publ. Astron. Soc. Pacif.* **111**, 50–56 (1999).
6. Latham, D. W., Stefanik, R. P., Mazeh, T., Mayor, M. & Burki, G. *Nature* **339**, 38–40 (1989).
7. Marcy, G. W. & Butler, R. P. *Astrophys. J.* **464**, L147–L151 (1996).
8. Marcy, G. W., Butler, R. P., Vogt, S. S., Fischer, D. A. & Liu, M. C. *Astrophys. J.* **520**, 239–247 (1999).
9. Marcy, G. W., Butler, R. P. & Fischer, D. A. *Bull. Am. Astron. Soc.* **194**, 14.02 (1999).
10. Noyes, R. W. *et al.* *Astrophys. J.* **487**, L111–L114 (1997).
11. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
12. Cameron, A. C., Horne, K., Penny, A. & James, D. *Nature* **402**, 751–755 (1999).
13. Guillot, T., Burrows, A., Hubbard, W. B., Lunine, J. I. & Saumon, D. *Astrophys. J.* **459**, L35–L38 (1996).
14. Goukenleuque, C., Bezaud, B., Jorget, B., Lellouch, E. & Freedman, R. *Astron. Astrophys.* (submitted).
15. Marley, M. S., Gelino, C., Stephens, D., Lunine, J. I. & Freedman, R. *Astrophys. J.* **513**, 879–893 (1999).
16. Seager, S. & Sasselov, D. D. *Astrophys. J.* **502**, L157–L161 (1998).
17. Sudarsky, D., Burrows, A. & Pinto, P. <http://xxx.lanl.gov/abs/astro-ph/9910504>
18. Charbonneau, D. *et al.* *Astrophys. J.* **522**, L145–L148 (1999).
19. Charbonneau, D., Brown, T. M., Latham, D. W. & Mayor, M. *Astrophys. J. Lett.* (in the press).
20. Henry, G. W., Marcy, G. W., Butler, R. P. & Vogt, S. S. *Astrophys. J. Lett.* (in the press).

Population biology

The prodigal fish

Stephen R. Palumbi

When Captain Ahab left harbour to hunt the white whale, he did not know exactly where to start. Cruising the hard sea of the nineteenth-century whaling fleets, he hunted and waited, primed for the sight of his unpredictable prey. Meeting another whale ship he would shout his frustration “Hast seen the white whale?”, thereby becoming one of the only North American males willing to ask directions.

Such is the fate of many ocean hunters—even those who drag the seas for real fish in modern commercial fleets. Understanding movement in the sea is one of the major unsolved problems of biological oceanography. In this issue, papers by Swearer *et al.*¹ (page 799) and Jones *et al.*² (page 802) take two large steps towards answering these questions. With different experimental approaches, they show that the larvae of reef fish, poor swimmers cast for weeks into the wide sea, are not always washed afar by ocean currents. Instead, some are retained near where they are spawned, and settle back onto the island reefs that their parents inhabited.

The secret life of fish has always held a geographical mystery. Most fish develop from planktonic larvae that drift for weeks or months in ocean currents. Diluted by the ocean's vast expanse, these larval fish have the potential to be transported thousands of kilometres, but whether they stay at home to

produce the next crop of juveniles, or drift off to mature far away, is generally unknown. This information is critical to fisheries models and coastal-management plans, and the basic assumption has been that larvae drift the oceans like vertebrate clouds, travelling vast distances and seldom going home.

Swearer *et al.*¹ attack this question by showing that the larvae of a common coral-reef fish, the Caribbean bluehead wrasse (*Thalassoma bifasciatum*), often spend their planktonic lives close to shore. They base their conclusions on tiny chemical tags that accumulate inside the otolith—a calcareous ossicle in a fish's ear.

Daily growth rings in an otolith can be counted to provide an estimate of the age at

which the fish left the larval stage³. Moreover, during otolith growth, trace minerals from the environment are included in the calcareous matrix. As a result each otolith is a chemical autobiography of a larva's life, recording which trace elements were available as the planktonic drifter grew and developed in the changing chemical sea^{4,5}.

Swearer *et al.* measured the elemental content of otoliths collected from settling wrasse larvae in the US Virgin Islands to estimate where larvae had drifted. Near-shore water masses contain higher levels of manganese, copper and barium than tropical, open-ocean waters. The otolith analysis showed that some larvae had high concentrations of these metals, as well as fast growth and large size at the transition from larva to adult. Other larvae had low metal content, slow growth and small size. The first syndrome is probably characteristic of larvae that were retained near shore and grew quickly on the thick planktonic soup that near-shore waters nurture. The second signature indicates larvae that were swept out to sea, but won the plankton lottery by being wafted back to shore at metamorphosis time. The big surprise in these results is that so many larvae—up to 50% in some of Swearer and colleagues' samples—had ‘retention signatures’. These larvae developed without a long open-ocean voyage, and so must have settled on the reefs of their natal island.

Jones *et al.*² took a more hands-on approach, and spent three months laboriously tagging ten million fish larvae around Lizard Island on the Great Barrier Reef. They labelled larvae by placing nearly two thousand nests of the damselfish *Pomacentrus amboinensis* in a dilute solution of tetracycline, allowing the developing eggs to absorb the fluorescent dye. The tetracycline was incorporated into the calcareous matrix of larval otoliths, and could be seen as a glowing ring under ultraviolet light.

After the ten million larvae hatched and disappeared, Jones *et al.* deployed light traps, floating like neon jellyfish in a night-time sea, to collect larvae at the end of their three-week planktonic period (Fig. 1). They caught 7,327 juvenile damselfish, and the otoliths from 5,000 of them were carefully dissected



Figure 1 On the Great Barrier Reef, many larvae of the damselfish *Pomacentrus amboinensis* stay near their point of origin. Jones *et al.*² found that they become juveniles on the reef that hatched them, despite a three-week period drifting in the plankton. (Picture courtesy of G. Jones.)

and viewed with an ultraviolet microscope. It took a month to find the first labelled otolith, glowing spectacularly under the microscope light, providing the first objective evidence that planktonic fish larvae can settle back onto the same reefs from which they were spawned.

All in all, 15 of the labelled larvae were collected — about one in 330. This figure seems small until compared with the larval output of the reefs of Lizard Island. Damselfish on Lizard Island produced one to two billion larvae during the tagging effort. Thus, Jones and colleagues tagged about 0.5–1% of larvae. Because 0.3% of the captured larvae showed the tetracycline tag, a substantial proportion (between one-third and two-thirds) of settling fish were estimated to be derived from eggs hatched on Lizard Island, with the rest being imported from elsewhere. This is a surprisingly high value for retention of larvae on a local reef, and suggests that many larvae stay close to their natal site for the entire planktonic period.

These two studies refute the conventional wisdom that planktonic dispersal must be widespread, and echo increasing evidence that oceanic conditions might favour retention. Oceanic gyres and eddies a few kilometres across can gather larvae in vortex centres, transporting them slowly as the centre of the gyre moves⁶. Simulations of ocean flows often show retention of particles near the point of release, or show surprisingly little movement of simulated larvae along coasts⁷. Genetic signatures of population separation can be exceedingly sharp. There is a steep cline — a gradient in gene frequency — in American oysters that spans only 20 km and has been geographically stable for over a decade⁸.

If long-distance marine dispersal is ecologically rare, this has profound implications

for the management of global fisheries and the maintenance of biological diversity in the sea. Local populations of fish may require management on a much finer scale than was previously thought. In addition, attempts to implement networks of marine protected areas to improve fisheries or maintain regional biodiversity rely on assumptions about ecological connectivity among populations⁹. The sizes of successful marine protected areas need to be scaled to average larval dispersal distance, and the distance between replicated beads in a reserve necklace may need to take larval dispersal distance into account.

These results suggest that more direct measurements of dispersal will be crucial in refining the management of marine ecosystems. Understanding how often fish larvae are retained on reefs, and whether these results also apply to continuous coastlines, with consistent patterns in the currents that run along the shore, will require complementary studies of other species and other oceanic settings. But for the first time, we know the fish come home. ■

Stephen R. Palumbi is in the Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. e-mail: spalumbi@oeb.harvard.edu

1. Swearer, S. E., Caselle, J. E., Lea, D. W. & Warner, R. R. *Nature* **402**, 799–802 (1999).
2. Jones, G. P., Milicich, M. J., Emslie, M. J. & Lunow, C. *Nature* **402**, 802–804 (1999).
3. Lagardere, F. & Troadec, H. *Mar. Ecol. Progr. Ser.* **155**, 223–237 (1997).
4. Thorrold, S. R., Jones, C. M. & Campana, S. E. *Limnol. Oceanogr.* **42**, 102–111 (1997).
5. Dufour, V., Pierre, C. & Rancher, J. *Coral Reefs* **17**, 23–28 (1998).
6. Lee, T., Clarke, M. E., Williams, E., Szmant, A. F. & Berger, T. *Bull. Mar. Sci.* **54**, 621–646 (1994).
7. Black, K. P. *Coral Reefs* **12**, 43–53 (1993).
8. Hare, M. P. & Avise, J. C. *Evolution* **50**, 2305–2315 (1996).
9. Roberts, C. M. *Science* **278**, 1454–1457 (1997).

Materials science

Injecting spin into electronics

Michael Oestreich

The operation of traditional semiconductor devices depends on the transport and storage of electronic charge. But electrons have spin as well as charge, and taking advantage of the spin could revolutionize electronics, leading to new devices such as spin transistors¹, spin memory storage or even spin quantum computers². One requirement for such devices is the efficient injection of spin-polarized carriers (electrons, or their positively charged counterparts, holes) into a semiconductor. This has proved to be a huge challenge. Now two groups have independently achieved success with real electrical devices. On pages 787 and

790 of this issue, Fiederling *et al.*³ and Ohno *et al.*⁴ report the injection of spin-polarized electrons or holes into a light-emitting diode, with efficiencies of about 90% and 2%, respectively. In both cases the diodes emit circularly polarized light, confirming the spin polarization of the carriers. So far these devices only work at low temperatures, but the race towards commercial semiconductor spin electronics is on.

Why is spin electronics in semiconductors interesting? One motivation comes from the discovery that alternating layers of non-magnetic materials and ferromagnetic metals (such as iron) show large changes in

their electrical resistance when a small magnetic field is applied. This change, known as giant magnetoresistance, can be greater than 60% at room temperature and results from the specific electron-spin orientation across the layers^{5,6}. The effect is already exploited in many applications, such as in the millions of reader heads in high-density magnetic computer disks. Another example, which might have a huge economic impact in the future, is nonvolatile magnetic random access memory that can retain information when switched off⁷. The use of spin effects in metals has led rapidly to commercial applications. Even more enticing is the dream of spin electronics in semiconductors, where you can easily engineer the electronic properties, combine conventional and spin electronics on one chip, and build optoelectronic spin devices.

Spin injection is currently the biggest obstacle facing semiconductor spin electronics. Why is this? Many groups have tried to use ferromagnetic metal contacts to inject spin-polarized electrons directly into a non-magnetic semiconductor. This seems like a good idea because electrons in ferromagnetic metals already have a preferential spin orientation at room temperature, even without an external magnetic field. But so far experiments have shown little or no spin polarization of carriers injected from a ferromagnetic metal into a semiconductor⁸. Dead magnetic layers at the metal–semiconductor interface and the huge difference in the number and energy of carriers are possible causes.

At the interface between a magnetic and a non-magnetic semiconductor these problems are absent. This is good news because, as yet, spin electronics has produced all-metal devices or devices combining metals and semiconductors. The goal is to make an all-semiconductor device, which would make it easier to incorporate spin electronics with traditional semiconductor technology. The two experiments reported in this issue^{3,4} show clearly that unpolarized carriers can become strongly spin polarized in a magnetic semiconductor and then electrically injected (that is, voltage driven) into a non-magnetic semiconductor with their spin orientation intact. Is the problem of spin injection in semiconductors finally solved?

Fiederling *et al.*³ show that voltage-driven spin injection from a semimagnetic to a non-magnetic semiconductor is highly efficient: the device injects 90% spin-polarized current into a light-emitting diode, which is good enough for industrial spin devices. They use an unusual semimagnetic semiconductor (a group II–VI compound; BeMnZnSe) as the 'spin aligner' to polarize the electrons before they are injected into the non-magnetic semiconductor (GaAs). Because the GaAs device is configured as a light-emitting diode, the polarized current



100 YEARS AGO

The recent researches of bacteriologists into the rôle played by insects as carriers of infection, and the hunt after microbes to locate their natural habitat, is a necessary procedure before it becomes possible to enter on a scientific crusade against them. In those diseases which may be caused by infection carried by insects, it is a more hopeful task to deal with the insects which we can see, than to deal with the microbe which lurks unseen and unheeded. ... The filariae, according to Manson, go through changes in the stomach of the mosquito, and finally make their way out into the water in which the insects have died, and man becomes infected by drinking the water. In this case and in Malaria the insect seems to act as an intermediary host to man. The mosquito — of which one species, the anopheles, seems to be mostly concerned — takes up the organism in the blood of the malarious subject, and, according to Manson, infects soil or water by dying in it; Ross and others, however, say it infects healthy persons by biting them after biting a malarious patient. It is interesting to note that most of our previous notions as to the localities and time of year that malaria occurs, and the precautions adopted to prevent being attacked still hold good, *mutatis mutandis*, for the mosquito theory.

From *Nature* 14 December 1899.

50 YEARS AGO

There are certain sections of organic chemistry of such a specialized nature that only a few persons have a real mastery of their intricacies. Examples are the chemistry and technology of azo-dyes, or of anthraquinone derivatives and related polycyclics. Certainly the chemistry, biochemistry and physiology of the sterinoids cannot be comprehended fully except by those who devote undivided attention to these subjects. ... Although cholesterol is a constituent of all animal cells, there was no inkling of its physiological importance when Windaus began his investigations in 1903, and this was also true in 1912, the approximate date of commencement of Wieland's work on related bile acids. The molecular structures were studied as interesting organic chemical problems, curiosities attractive because of their difficulty and the unique character of the group.

From *Nature* 17 December 1949.

leads to polarized light emission. The spin aligner works well at low temperatures and provides a new technique for probing spin transport in semiconductors. But it has been known for some time that this material is only semimagnetic at temperatures much below liquid nitrogen (a few kelvin). It is probably not possible to improve this temperature range by increasing the magnetic Mn concentration, because Mn ions usually couple antiferromagnetically in this system at high concentrations. A ferromagnetic II–VI semiconductor working as a spin aligner at room temperature would therefore be a great surprise.

Ohno *et al.*⁴ take a different approach. They use GaMnAs (a group III–V semiconductor) as a spin aligner for positively charged holes (rather than electrons). Spin-polarized holes are electrically injected from the GaMnAs, through a GaAs spacer, into a quantum well configured as a light-emitting device. At the same time unpolarized electrons are injected into the quantum well, where they recombine with the holes, emitting polarized light. GaMnAs has a huge advantage over II–VI semiconductors because it can be ferromagnetic at moderate temperatures and perhaps even at room temperature if the Mn concentration can be increased. Moreover, spin injection in such ferromagnetic semiconductors does not require an external magnetic field.

The polarization of the holes in the GaMnAs spin aligner is probably close to

100%, similar to the electron spin polarization seen by Fiederling *et al.*. But the polarization of the emitted light, which is used to probe the spin of the injected carrier, is more than ten times lower in Ohno and colleagues' device. The reasons are not yet clear but two explanations should be considered. First, the optical selection rules for heavy-hole transitions in ideal quantum wells prohibit circularly polarized light emission in certain directions (normal to the growth direction), even if all the hole spins are polarized. The correlation between carrier spin and polarization of the emitted light is therefore not as straightforward as in Fiederling and colleagues' device. Second, the orientation of the hole spin in bulk GaAs is, unlike the orientation of the electron spin, extremely unstable, and even in quantum wells the hole-spin orientation disappears incredibly quickly at moderate hole temperatures^{9,10}. Therefore, hole-spin injection will probably not be important for future spin electronics.

Despite these limitations, the two experiments represent real progress in spin electronics and they directly point the way to commercial spin-sensitive semiconductor devices. Any ferromagnetic semiconductor that injects electrons and can operate at room temperature will solve the problem, and so the search goes on.

Michael Oestreich is in the Physics Department, University of Marburg, Renthof 5, D-35032 Marburg, Germany.
e-mail: Michael.Oestreich@physik.uni-marburg.de

Animal behaviour

The male tail

Differences in the size of males and females are common in nature, and the simple shape of snakes makes them excellent subjects for studies of the phenomenon. In many species, males have longer tails than females, and this dimorphism seems to have arisen early in snake evolution. But why? From studies in Manitoba, Canada, R. Shine and colleagues provide evidence that long-tailed males make better lovers than their shorter rivals (*Proc. R. Soc. Lond. B* 266, 2147–2151; 1999).

Mating of red-sided gartersnakes (*Thamnophis sirtalis parietalis*) takes place just after they emerge from over-wintering, when they form mating balls (see picture). Shine *et al.* found, first, that longer-tailed males tended to



be more successful in copulating. Second, they took advantage of an event in which, due to the enthusiasm of males to get in on the act, over 100 had suffocated at the bottom of a huge ball. From these dead specimens, the authors found that the length of the tail and the size of a male's reproductive organ, the hemipenis, are correlated. Finally, snakes with shortened tails, due to predation or

misadventure, obtained far fewer matings than their intact compatriots.

These results are only correlations, and it is unclear how natural selection has influenced the evolution of this trait. Females may select mates using tail length as an indicator of sexual potency, or perhaps a longer tail makes it easier for males to hold their partner in a loving embrace.

Jeff Harvey

1. Datta, S. & Das, B. *Appl. Phys. Lett.* **56**, 665–667 (1990).
2. Imamoglu, A. *et al. Phys. Rev. Lett.* **83**, 4204–4207 (1999).
3. Fiedlerling, R. *et al. Nature* **402**, 787–790 (1999).
4. Ohno, Y. *et al. Nature* **402**, 790–792 (1999).
5. Grünberg, P., Schreiber, R., Pang, Y., Brodsky, M. B. & Sowers, H. *Phys. Rev. Lett.* **57**, 2442–2445 (1986).
6. Schad, R. *et al. Phys. Rev. B* **59**, 1242–1248 (1999).
7. Parkin, S. S. P. *et al. J. Appl. Phys.* **85**, 5828–5833 (1999).
8. Hammar, P. R., Bennet, B. R., Yang, M. J. & Johnson, M. *Phys. Rev. Lett.* **83**, 203–206 (1999).
9. Meier, F. & Zakharchenya, B. P. (eds.) *Optical Orientation* (North-Holland, Amsterdam, 1984).
10. Marie, X. *et al. Phys. Rev. B* **60**, 5811–5817 (1999).

Spongiform encephalopathies

Breech-birth prions

James Hope

Creutzfeldt–Jakob disease, Gerstmann–Straussler–Scheinker (GSS) syndrome, scrapie and bovine spongiform encephalopathy are part of a family of neurodegenerative diseases caused by changes in the metabolism of a plasma-membrane protein, PrP^C or prion protein. These prion diseases can be acquired by infection, arise sporadically by an unknown mechanism, or result from inheritance of a defective *PrP* gene. An abnormal PrP (PrP^{Sc}) is characteristic of most prion diseases, but in some inherited cases the PrP that accumulates in the brain is not PrP^{Sc} but ^{Ctm}PrP, a transmembrane form of the protein¹. And on page 822 of this issue, Hegde and co-workers² show that ^{Ctm}PrP may be a neurotoxic molecule that is common to both the genetic and the acquired forms of prion diseases.

The prion hypothesis holds that an abnormal prion protein (PrP^{Sc}) may be the sole component of the ‘transmissible particle’ that causes acquired forms of disease. But there are several anomalies. For example, high levels of a GSS-syndrome mutant PrP (P101L) in the brains of transgenic mice (the Tg[MoPrP(P101L)] strain) induced a spon-

taneous neurological disorder in the absence of conventional PrP^{Sc}, although this disorder could be transmitted to a limited range of host rodents³. A subset of people with GSS syndrome has been found with a different mutation (A117V) in the *PrP* gene that also does not produce conventional PrP^{Sc}. However, this form of the disease cannot be transmitted to laboratory rodents⁴. Further anomalies in the prion hypothesis have been exposed by cross-species transmission experiments^{5,6}, in which recipient mice develop disease without the formation of recognizable PrP^{Sc}.

These anomalies have led some to reject the prion hypothesis in favour of a more conventional viral agent as the cause of transmissible spongiform encephalopathies. Hegde and co-workers², by contrast, have looked for a solution to these difficulties by understanding the effect of mutations on the cell-membrane topography and stability of PrP^C, *in vitro* and *in vivo*. Normally, PrP^C is fully translocated across the membrane of the endoplasmic reticulum, then anchors to its luminal face before being processed and transported to the cell surface. But some-

times it can get trapped in the membrane, and it forms the ^{Ctm}PrP form. The carboxy-terminal (Ctm) segment of this protease-sensitive intermediate pokes out into the lumen of the endoplasmic reticulum, whereas its amino-terminal domain is stuck in the cytosol on the other side — like a breech baby, it seems to be born tail first.

In a previous publication¹, Hegde and colleagues found a correlation between the propensity of mutants to form ^{Ctm}PrP and their ability to induce neurodegenerative disease (in a dose-related manner) in transgenic mice. Elevated levels of ^{Ctm}PrP were also associated with a case of GSS syndrome (A117V), and ^{Ctm}PrP was suggested to be involved in the brain degeneration associated with non-transmissible, ‘genetic’ prion diseases. Hegde *et al.*² have now tested whether this ^{Ctm}PrP-associated disease can be transmitted. To do this, they took brain homogenates from one line of transgenic mice, known as [Tg[SHaPrP(KH→II)_H]], that express high levels of a hamster ^{Ctm}PrP. These mice rapidly develop spontaneous disease, characterized by high levels of ^{Ctm}PrP but no detectable PrP^{Sc}. The authors then inoculated a panel of transgenic mice and hamsters with these brain homogenates. They observed no significant differences in survival times between animals inoculated with Tg[SHaPrP(KH→II)_H] mouse brain and those inoculated with control tissues. In other words, ^{Ctm}PrP-associated prion disease is not transmissible.

The hamster PrP protein expressed in Tg[SHaPrP(KH→II)_H] mice has amino-acid changes at codons 109 (where lysine is replaced by isoleucine) and 110 (histidine by isoleucine). Hegde *et al.* next identified other mutations leading to an *in vitro* increase in levels of ^{Ctm}PrP, and produced lines of transgenic mice expressing high and low levels of the equivalent PrP proteins. High expressors invariably developed spontaneous disease, with ^{Ctm}PrP but no protease-resistant PrP^{Sc} in their brains. Low expressors were clinically normal, and did not show accumulation of ^{Ctm}PrP. The authors defined the product of the expression levels in Tg[SHaPrP(KH→II)_H] mice, and the ratio of ^{Ctm}PrP to total PrP formed in a coupled transcription–translation system, as the ‘Ctm-index’ — a semi-quantitative measure of the propensity of a transgenic line to generate ^{Ctm}PrP. They then used their transgenic mice to see whether there is a relationship between ^{Ctm}PrP and PrP^{Sc} during acquired disease.

When Hegde and co-workers infected their mouse lines with a hamster scrapie strain (Sc237), they found that the amount of PrP^{Sc} accumulating at the onset of clinical disease seemed to be inversely related to the Ctm-index. This was the case irrespective of whether comparisons were made between mice with constant levels of PrP expression

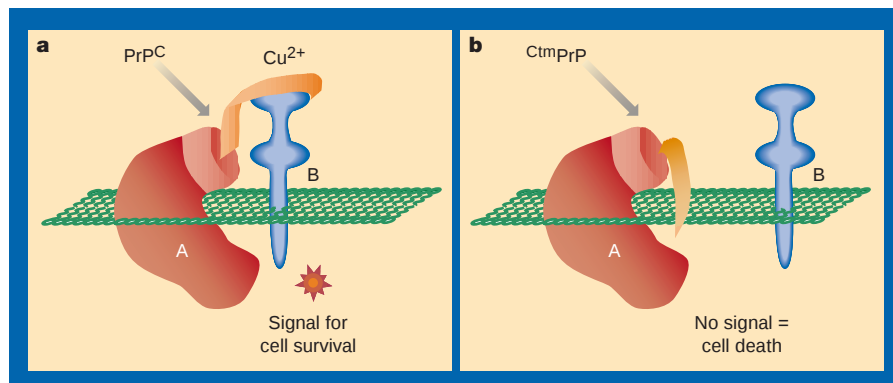


Figure 1 Possible role of ^{Ctm}PrP in cell death. a, The normal, full-length prion protein (PrP^C) may act as a co-receptor on the cell surface, facilitating the juxtaposition of two cell-surface transmembrane molecules (A and B). This may generate a signal for cell survival in the cytosol. This process does not require PrP^C, although PrP^C could act as a sensor to the extracellular environment (detecting, for example, levels of Cu²⁺)⁹ via its amino-terminal sequence, providing an extra level of cellular recognition. b, Failure of ^{Ctm}PrP to bind B could induce cell death by disrupting the association of A and B. This mechanism may also account for the lethal (but recessive) effects of expression of an amino-terminally truncated PrP (ref. 7) and the *Dpl* gene product⁸. The laminin receptor protein¹⁰ and heparan sulphate proteoglycans^{11,12}, possibly those of the syndecan family, have been shown to bind PrP^C and are candidates for A and B, respectively.

but different mutations, or between mice with the same mutation and different levels of expression. Amino-terminally truncated PrP (ref. 7) or Dpl (a PrP-like molecule)⁸, both of which lack an amino-terminal repetitive sequence, have been linked to neurodegenerative diseases, and ^{Ctm}PrP may induce cell death by a similar mechanism (Fig. 1).

All the available data fit with a model for an interrelationship of ^{Ctm}PrP and PrP^{Sc}, in which ^{Ctm}PrP causes the neurodegeneration seen in both genetic and transmissible prion protein diseases. One prediction of the model is that the accumulation of PrP^{Sc} should increase the generation of ^{Ctm}PrP. The authors used an ingenious approach to test this prediction and solve the problem of discerning PrP^{Sc} from ^{Ctm}PrP in infected transgenic mice. They produced a mouse line that makes both mouse PrP (MoPrP) and hamster PrP (SHaPrP), and infected that line with the RML strain of scrapie. This strain can convert only mouse PrP — not hamster PrP — to PrP^{Sc}. Using a hamster-specific monoclonal antibody to distinguish between ^{Ctm}SHaPrP and MoPrP^{Sc}, they showed a positive correlation between the amounts of PrP^{Sc} and ^{Ctm}PrP accumulating in the brain during infection.

When some prion diseases are transmitted between species, PrP^{Sc} is not observed. Could the existence of a neurotoxic ^{Ctm}PrP resolve this anomaly? Yes if one assumes that, when transmitting across species barriers (such as from cow to mouse^{5,6} or from human to Tg[MoPrP(P101L)] mouse⁶), the balance between ^{Ctm}MoPrP and MoPrP^{Sc} production is heavily in favour of ^{Ctm}MoPrP. This would lead to a transmitted, neurological disease provoked by ^{Ctm}PrP accumulation in the virtual absence of PrP^{Sc}. Similarly, if Tg[MoPrP(P101L)] mice³ have a high Ctm-index, the anomaly of their spontaneous disease in the absence of PrP^{Sc} could be explained by the preferential accumulation of ^{Ctm}PrP. However, it then becomes an even greater mystery as to why this MoPrP(P101L) 'spontaneous' disease should be transmissible whereas the other PrP spontaneous diseases are not. ■

James Hope is at the Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire RG20 7NN, UK.

e-mail: james.hope@bbsrc.ac.uk

- Hegde, R. S. *et al.* *Science* **279**, 827–834 (1998).
- Hegde, R. S. *et al.* *Nature* **402**, 822–826 (1999).
- Hsiao, K. K. *et al.* *Proc. Natl Acad. Sci. USA* **91**, 9126–9130 (1994).
- Tateishi, J. *et al.* *Neurology* **40**, 1578–1581 (1990).
- Lasmezas, C. I. *et al.* *Science* **275**, 402–405 (1997).
- Manson, J. *et al.* *EMBO J.* **18**, 6855–6864 (1999).
- Shmerling, D. *et al.* *Cell* **93**, 203–214 (1998).
- Moore, R. C. *et al.* *J. Mol. Biol.* **292**, 797–817 (1999).
- Brown, D. R. *et al.* *Nature* **390**, 684–687 (1997).
- Rieger, R. *et al.* *Nature Med.* **3**, 1383–1388 (1997).
- Caughey, B. *et al.* *J. Virol.* **68**, 2135–2141 (1994).
- Harris, D. A. *et al.* *Curr. Top. Microbiol. Immunol.* **207**, 77–93 (1996).

Earth science

The mantle's lava lamp

Marcia McNutt

Almost all geological observations at Earth's surface can be explained by the relative drift and interaction of tectonic plates. The two exceptions are hotspots (isolated points of high melt delivery to Earth's surface, such as Hawaii) and superswells (clusters of midplate hotspots on very broad topographical domes, such as French Polynesia). Hotspots and superswells offer the best clues as to the dynamic state of Earth's mantle. As described by Davaille on page 756 of this issue¹, their essential features have for the first time been reproduced in laboratory experiments of convection scaled to Earth-like conditions.

Previous numerical and laboratory simulations of mantle convection have failed to re-create the broad (several thousands of kilometres), oscillatory upwellings neces-

sary to produce superswells, as determined from modelling of geophysical data². Narrow plume-type upwellings that could explain hotspots are readily produced in models with sharp temperature gradients separating convecting layers of different density, the most dramatic of which must occur at the core–mantle boundary. But plumes originating at the core–mantle boundary should be much hotter than the temperatures inferred from geochemical analyses of hotspot lavas³.

The key to Davaille's success was in recognizing that even very subtle density layering (around 1%) within the mantle itself would have important geodynamic consequences. In her laboratory tank tests, Davaille observed the modes of convection that developed in viscous fluids stratified in den-

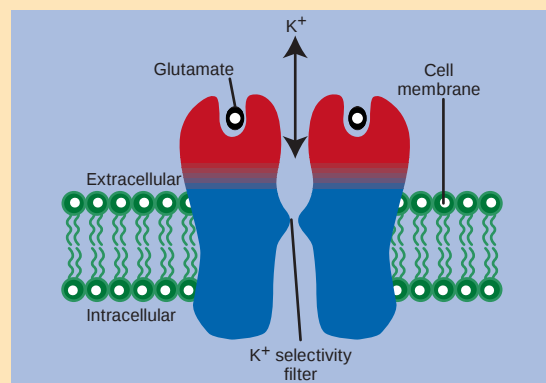
Ion channels

Common origins

Glutamate receptors and potassium channels both allow the passage of charged ions between the inside and outside of cells, but are mechanistically quite different. So the discovery in a bacterium of what looks like a 'hybrid' between them, as described by G.-Q. Chen and colleagues elsewhere in this issue (*Nature* **402**, 817–821; 1999), is quite an event.

Glutamate receptors are permeable to several different ions and are opened, or 'gated', by their ligand, glutamate. By contrast, as their name suggests, potassium channels show exquisite selectivity for potassium ions. These channels can be gated either by intracellular ligands, or by voltage (a mechanism that is coincidentally investigated in two other papers in this issue, on pages 809 and 813).

Chen *et al.* have identified a glutamate receptor, GluR0, from the cyanobacterium *Synechocystis*. This is the first prokaryotic glutamate receptor to be described, and a surprise in itself. But remarkably the



authors also reveal that it is endowed with a potassium-selective pore.

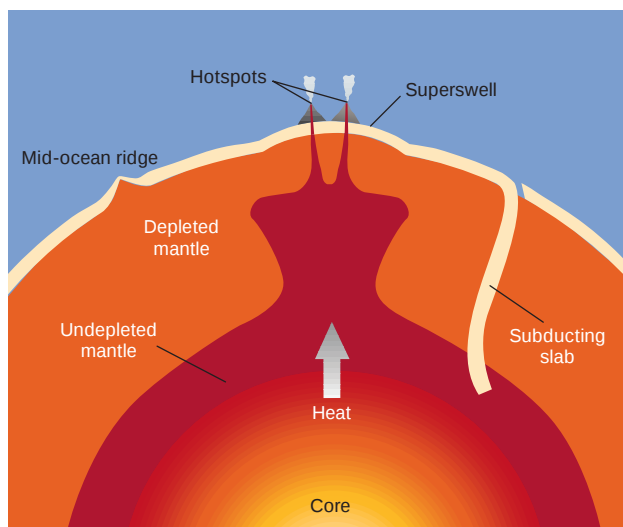
The extracellular regions of GluR0 show sequence similarity to the ligand-binding domains of glutamate receptors (red in the figure here) and the pore region looks like an inverted potassium-channel pore (blue). The exciting implication is that potassium channels and glutamate receptors may have originated from a common ancestor, and so may have common architectural and functional features.

Last year, the first potassium-channel structure

was solved by D. A. Doyle *et al.* (*Science* **280**, 69–77; 1998). The channel, known as KcsA, is found in the bacterium *Streptomyces lividans*, and is gated by an intracellular ligand. The structural information allowed predictions to be made about the structure of other potassium-channel pores. Until now, however, there has been no such information available for glutamate receptors. A further aspect of the work by Chen *et al.* is that they show that the KcsA structure can be used to make predictions about glutamate-receptor pores as well.

Lesley Anson

Figure 1 Thermochemical convection in Earth's mantle as it might be now, inferred from Davaille's laboratory simulations¹ with a buoyancy ratio of <0.5 . Heat from the lower mantle and core causes oscillatory doming of an intrinsically denser layer of mantle silicates just above the core. The deeper layer provides a more primitive source of hotspot lavas because, unlike the upper mantle, this region has not lost its volatiles and easily melted radioactive elements during previous plate-tectonic melting cycles.



sity, heated from below, and cooled from above. Results of the experiments depended on the ratio of the intrinsic chemical density difference between the layers to the density anomaly caused by thermal expansion of heated fluid. When this ratio is greater than 1, heating of denser fluid from below does not generate enough thermal buoyancy to offset the chemical stratification, except locally in isolated plumes that rise from the density interface to the surface.

On the other hand, when the buoyancy ratio is between about 0.3 and 0.5, the interface between the stratified fluids deforms into broad domes that rise to the top of the tank, cool and sink back down because of their greater chemical density. This process, which is akin to that seen in a lava lamp⁴, produces an oscillatory pattern of superswell-type features reminiscent of the 100-million-year cycle of superswell formation in the central Pacific. For certain ratios of viscosity contrast between the upper and lower layers, relatively cool plumes rise from the thin thermal boundary layer at the upper surface of the domes (Fig. 1).

Davaille suggests that Earth may have originally formed in the layered-convection regime (buoyancy ratio >1), but that gradual mixing of the two layers by plume-type upwellings and sheet-type downwellings attenuated the chemical layering to the point that the buoyancy ratio is now less than 0.5. (It is in this sense that the lava-lamp analogy to her model breaks down. Davaille's fluids are completely miscible, unlike those of a lava lamp.) Furthermore, her simulations suggest that the oscillatory doming of superswells rapidly homogenizes the mantle zonation so that there are no more than five cycles of superswell formation. Given evidence from the Pacific that the timescales for superswell rise and fall are of the order of 100,000 years, this would mean that in much less than a billion years Earth will evolve to a well-mixed state of whole-mantle convection.

This model may help to reconcile several conflicting observations as to the extent of layering in mantle convection. The weak stratification thought to represent the present-day situation is insufficient to prevent cold lithospheric slabs, descending at the ocean trenches, from penetrating to depths of 1,000 km or more, as has now been conclusively imaged with seismic tomography⁵. If chemical stratification was indeed stronger earlier in Earth history, it would facilitate the preservation of a deep source for hotspot lavas, more enriched in volatiles and radiogenic isotopes as compared with the mid-ocean-ridge lavas derived from the uppermost mantle⁶.

Of course, much work remains to be done before it will be possible to declare vic-

tory for a model of mantle mixing that is consistent with all geochemical and geophysical observations. Davaille's experiments were conducted in a rectangular tank, not a spherical shell. Her fluids, unlike mantle materials, do not exhibit temperature-dependent variations in viscosity (which influence mixing times) nor realistic depth variations in physical properties, such as the coefficient of thermal expansion. Perhaps most importantly, the lack of temperature-dependent viscosity precludes the formation of more rigid plates in a surface thermal boundary layer, the recycling of which in subduction zones is an important mechanism for reintroducing chemical heterogeneity into the real Earth.

Nevertheless, these experiments point the way for future laboratory and numerical modelling. It can no longer be assumed that chemical density heterogeneities of the order of 1–2% are geodynamically inconsequential. And a snapshot of mantle convection today (as imaged by seismic tomography, for example) may not be representative of its geodynamic behaviour throughout all four billion years of Earth history.

Marcia McNutt is at the Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA.

e-mail: mcnutt@mbari.org

1. Davaille, A. *Nature* **402**, 756–760 (1999).
2. McNutt, M. *Rev. Geophys.* **36**, 211–244 (1998).
3. Albers, M. & Christensen, U. R. *Geophys. Res. Lett.* **23**, 3567–3570 (1996).
4. www.steeles.com/lavalamps/
5. Van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. *Nature* **386**, 578–584 (1997).
6. Zindler, A. & Hart, S. *Annu. Rev. Earth Planet. Sci.* **14**, 493–571 (1986).

Lifespan

The effects of sensory deprivation

James H. Thomas

Once, lifespan genetics was largely the domain of theorists, who tried to explain why an organism's genes so cavalierly allow individual somas to die. But a flood of papers on the nematode worm *Caenorhabditis elegans* has brought the subject into the realm of serious experimental analysis. The latest studies^{1,2}, including a report by Apfeld and Kenyon¹ on page 804 of this issue, indicate that the nervous system has a key function in regulating lifespan. Perhaps we are, indeed, only as old as we think we are.

Most of the genes known to regulate adult lifespan also affect formation of the developmentally distinct larval stage called the dauer larva. Formed in response to crowding and other stressors, the dauer larva is adapted to long-term survival under harsh conditions, including the absence of food. Unlike life-

span, dauer formation is very amenable to genetic analysis, and it has been the subject of many studies. Progress in understanding genetic control of lifespan in *C. elegans* began with the finding that mutations in the dauer-formation gene *daf-2* extend lifespan dramatically³. This finding allowed researchers to tap into existing information about the genetic pathways that control dauer formation and, more importantly, to use dauer phenotypes to analyse the genes that control lifespan.

This sudden change in the ease of analysis resulted in an explosion of progress, culminating in a fairly complete description of a lifespan-regulating transduction pathway headed by *daf-2*. This gene encodes an insulin-receptor-like protein⁴, and the steps that occur downstream of the DAF-2 receptor parallel the biochemically defined

insulin-signalling pathway in mammals. Specifically, activation of DAF-2 influences the rate of synthesis and degradation of a second messenger called phosphatidylinositol-1,4,5-trisphosphate, $\text{PtdIns}(1,4,5)\text{P}_3$ (refs 5,6). The resulting increase in the level of $\text{PtdIns}(1,4,5)\text{P}_3$ activates a protein-kinase cascade^{7,8}. One output of this insulin-signalling pathway, which was first discovered in *C. elegans*, occurs via the forkhead-like transcription factor DAF-16 (refs 9,10). Mutations that result in activation of DAF-16 lead to a dramatic extension of lifespan (and increased dauer formation), and loss of DAF-16 function prevents this extension.

One gap in our understanding has been what acts upstream of the DAF-2 insulin receptor to regulate its activity. The answer is presumably insulin-like ligands¹¹, but what regulates their synthesis and secretion? Dauer formation involves sensory transduction, and the roles of several specific ciliated sensory neurons in dauer formation have been described¹². Many genes have been identified in *C. elegans* that affect the structure and function of broad sets of sensory neurons, and Apfeld and Kenyon¹ now show that mutations in nearly all of these genes cause increased lifespan, although this effect is much weaker than that caused by mutations in *daf-2*. These sensory mutations do not seem to influence feeding rate, timing of development or fertility, suggesting that their effect on lifespan is relatively direct.

In *C. elegans*, lifespan can also be increased if the germ line is eliminated by laser microsurgery¹³. When Apfeld and Kenyon did this in the sensory-deficient mutants they found that the extension of lifespan was roughly additive, suggesting that the two pathways act independently. Finally, the authors showed that sensory-deficient mutations do not further extend the lifespan of worms mutated in the *daf-2* gene. Moreover, the extended-lifespan effect in sensory-deficient worms is partially suppressed by mutations in the *daf-16* gene. These results indicate that the extension of lifespan caused by the sensory defects results, at least in part, from effects on the insulin-signalling pathway.

What do these striking results mean for the normal process of ageing? The authors' interpretation is simple and sexy — they assert that environmental signals act through sensory neurons to control lifespan. Supporting this view is the fact that there are many insulin-like proteins in *C. elegans*¹¹. By analogy with vertebrate insulins, the worm insulins are presumably packaged into secretory vesicles for release from excitable cells. These proteins are a potential mechanistic link between sensory processes and the *daf-2* insulin-signalling pathway. It is possible that insulin release is regulated by sensory neurons that respond to environmental cues,

and that sensory defects reduce this release. The result is a partial failure in activation of the DAF-2 receptor.

But is the link between sensory processes and insulin signalling as direct as a failure to respond to specific sensory cues, as Apfeld and Kenyon suggest? There is reason to doubt this simple idea. In the sensory-deficient mutants, the affected sensory neurons don't just fail to function — they adopt physiologically abnormal states. In particular, the neurons that regulate dauer formation are in an abnormal state that influences dauer formation in ways not characteristic of the wild type¹⁴. For example, in these sensory mutants affecting the structure of cilia, at least one of the dauer-regulating neurons is thought to release the transforming growth factor- β -related protein DAF-7 constitutively¹². This activity accounts for the dauer-defective phenotype of these mutants. A morphological correlate to this abnormal state is the growth of aberrant axonal processes by many of these neurons¹⁵. Considering the complexity and variety of the defects, simple interpretations of these results would be incautious.

Nevertheless, Apfeld and Kenyon provide excellent evidence that physiologically abnormal sensory neurons can exert a strong influence on the lifespan of *C. elegans*. This evidence is consistent with a key function for these neurons in the related process of dauer formation, and with the regulation of insulin release by neurons. What it does not provide is specific information about what the normal sensory involvement might be. Discovering whether there are specific environmental modulators of lifespan and, if so, what they are and how they act, is a challenge for the future. ■

James H. Thomas is in the Department of Genetics, University of Washington, Seattle, Washington 98195, USA.

e-mail: jht@genetics.washington.edu

1. Apfeld, J. & Kenyon, C. *Nature* **402**, 804–809 (1999).
2. Ailion, M., Inoue, T., Weaver, C. I., Holdcraft, R. W. & Thomas, J. H. *Proc. Natl Acad. Sci. USA* **96**, 7394–7397 (1999).
3. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. *Nature* **366**, 461–464 (1993).
4. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *Science* **277**, 942–946 (1997).
5. Morris, J. Z., Tissenbaum, H. A. & Ruvkun, G. *Nature* **382**, 536–539 (1996).
6. Ogg, S. & Ruvkun, G. *Mol. Cell* **2**, 887–893 (1998).
7. Paradis, S., Ailion, M., Tokar, A., Thomas, J. H. & Ruvkun, G. *Genes Dev.* **13**, 1438–1452 (1999).
8. Paradis, S. & Ruvkun, G. *Genes Dev.* **12**, 2488–2498 (1998).
9. Ogg, S. *et al.* *Nature* **389**, 994–999 (1997).
10. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. *Science* **278**, 1319–1322 (1997).
11. Duret, L., Guex, N., Peitsch, M. C. & Bairoch, A. *Genome Res.* **8**, 348–353 (1998).
12. Schackwitz, W. S., Inoue, T. & Thomas, J. H. *Neuron* **17**, 719–728 (1996).
13. Hsin, H. & Kenyon, C. *Nature* **399**, 362–366 (1999).
14. Vowles, J. J. & Thomas, J. H. *Genetics* **130**, 105–123 (1992).
15. Peckol, E. L., Zallen, J. A., Yarrow, J. C. & Bargmann, C. I. *Development* **126**, 1891–1902 (1999).

Daedalus

Publication paradise

The worth of a scientific paper is judged by its citations in other papers. Daedalus's electronic publishing database of last week ranked the citations themselves by the citation score of the cited authors. He now has more consequences of pure Internet-based publishing.

For a start, it will relax all size constraints. Many scientific papers are jargon-packed and atrociously written because the authors have to squeeze their message into limited journal space. And thousands of interesting small observations are never published at all, being too brief for any paper. In particular, many negative results languish in file drawers. But an electronic system, with the total indexing it makes possible, could publish papers of any size. All would have the chance to be retrieved, studied and cited, bringing prestige to their authors. And with number of citations, not number of papers, as the criterion of prestige, size doesn't matter.

How would the system be paid for? A scientific paper is an advertisement for its authors (indeed, some are little else). So page charges levied on those authors should work well. The tricky problem is refereeing, to keep out (or at least point out) the nutters and incompetents. A democratic database should accept any paper, however mad, provided the author could pay for it. But he would be wise to offer it to a referee, who would 'mark' it, and append comments in the form of a separate note cited in the paper. This citation would neatly reward the referee for his work. The system would multiply the referee's mark, positive or negative, by his own citation score. An author who approached a high-ranking referee would thus be taking a big risk — condemnation would give the paper a very negative mark. Users could scan the database through a filter cutting out papers below a certain mark, thus excluding the low-ranked stuff.

The whole database would probably divide itself into self-contained mutually citing regions. Some would be inhabited by cliques of nutters all citing each other. Some would form rival schools of mutually negative refereeing. But science, alone among philosophies, converges on consensus. The biggest mutually citing region would be the domain of respectable, authoritative scientific opinion. **David Jones**

The further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Obituary

Johannes A. van Paradijs (1946–99)

Jan van Paradijs, who died in Amsterdam on 2 November, was one of the world's foremost astrophysicists. He will probably be remembered most for the discovery, in 1997, of the first optical afterglow of a γ -ray burst, which established the extragalactic nature of these bursts and solved what had for some 25 years been one of the greatest problems of astrophysics. Because of this first optical sighting, which has since been followed by a dozen others, we now know that γ -ray bursts occur in very distant galaxies and represent the largest explosions of energy in the Universe since the Big Bang.

Born into a bricklayer's family in Haarlem, The Netherlands, van Paradijs was the eldest of seven children. Thanks to the intervention of the headmaster of his Roman Catholic primary school, his parents sent him to a Haarlem secondary school that prepares students for higher education. In 1963 he began studying mathematics, physics and astronomy at the University of Amsterdam. He excelled in all three fields and graduated in 1970. During those years he played top-league basketball for his home town Haarlem. He also took many university classes in philosophy, and clear logic and consistent reasoning became second nature to him in all his activities. Later, he would teach these qualities, both in words and by example, to students and colleagues.

Van Paradijs carried out his PhD research on the structure and chemical composition of cool giant stars under the guidance of David Koelbloed. The knowledge he gained in stellar spectroscopy and photometry, as well as in the theory of stellar atmospheres, prepared him well for his later research on X-ray binaries and neutron stars. When I came to Amsterdam in 1974, he had just become a tenured research associate and decided to join me in studies of binary X-ray sources, complementing my theoretical work with careful observations.

X-ray binaries are the strongest sources of X-rays in the Galaxy, in which gas flows from a normal star to a nearby compact companion, such as a neutron star or black hole. Van Paradijs carried out systematic studies of the orbits and light curves (variation of light over time) of the normal stars in these systems, with the aim of accurately determining the mass of a companion neutron star. As their name suggests, neutron stars consist almost entirely of neutrons and so are very dense and small.



High-energy astrophysicist who hunted big astronomical game

In 1975 van Paradijs succeeded in measuring the first ever mass of a neutron star: Vela X-1. This neutron star is about twice as heavy as the Sun, but has a diameter of only about 20 km. He realized the great importance of these accurate mass determinations for fundamental physics. Together with radius measurements of neutron stars — a topic he addressed in later studies of X-ray burst sources — they are the only way to study nuclear matter at densities far beyond what can be reached in the laboratory.

From 1977 to 1979 van Paradijs worked at MIT, where he became interested in X-ray bursts — powerful, brief outbursts of X-rays that were discovered by a Dutch satellite in 1975. Colleagues at MIT, George Clark, Walter Lewin and Paul Joss, were using a small astronomy satellite to observe such X-ray bursts and made many important discoveries. His close friendship with Walter Lewin dated from this time — they produced over 120 papers together, out of 400 or so scientific papers van Paradijs produced during his lifetime.

A trait of van Paradijs' work was his talent for seeing the 'big picture' through a multitude of often confusing observational facts. A prime example was his demonstration that X-ray burst sources are neutron stars and not black holes as had been generally thought. He showed that the peak luminosities of many bursts were always the same, leading to the idea of a

'standard candle'. From this he was able to measure the diameters of several burst sources, and found them all to be about 16 km, typical of a neutron star. Among his other discoveries were the demonstration that X-ray-burst sources are members of binary systems, and the first spatially resolved spectroscopic mapping of an accretion disk that forms between two stars in a low-mass binary.

In 1978, while working at the European Southern Observatory in Chile, together with Holger Pedersen from Copenhagen, van Paradijs discovered the first optical flash from an X-ray burst, by simultaneously combining telescope and satellite data. Such simultaneous observations of unpredictable, brief events require much organization and a flair for logistics. There can be no doubt that van Paradijs' expertise in this area gave him a head start over other groups when in 1996 the BeppoSAX satellite began to provide accurate positions of γ -ray bursts. It was therefore no coincidence that, together with his students Paul Groot and Titus Galama, he was the first to detect the optical afterglow from a γ -ray burst and to establish its extragalactic nature.

Jan van Paradijs married the Greek astrophysicist Chryssa Kouveliotou in 1992. She works at NASA's Marshall Space Flight Center in Huntsville, Alabama, where from 1993 Jan was a part-time physics professor at the University of Alabama. He divided his time between Huntsville and Amsterdam, where he was also a professor. The marriage was a very happy one, each partner strongly supporting the other in research as well as in personal life. He was co-author on her 1998 paper reporting the discovery of magnetars — neutron stars with giant magnetic fields of 10^{15} Gauss — just as she was co-author on his paper reporting the discovery of the first γ -ray-burst afterglow. All in all, in the last eight years they published more than 90 papers together.

It came as a great shock that at the peak of their careers, Jan suddenly appeared to be gravely ill. During the past seven months Chryssa practically suspended her career to care for him until the end. Van Paradijs guided the research of over 25 PhD students in Amsterdam and Huntsville. He will be remembered by them all as an outstanding scientist and teacher, and a fine colleague and friend.

Edward van den Heuvel

*Edward van den Heuvel is at the Astronomical Institute "Anton Pannekoek" and the Center for High Energy Astrophysics, University of Amsterdam, Kruislaan 403, NL-1098 SJ Amsterdam, The Netherlands.
e-mail: edvdh@astro.uva.nl*

The future of cloning

J. B. Gurdon and Alan Colman

It is now possible to make clones, or exact genetic copies, of sheep, cows, goats, mice and, probably, humans. This opens the way towards the production of replacement body parts from adult cells.

Cloning techniques have been in use for centuries. The practice of taking cuttings is universal among gardeners, and large companies now propagate desirable plant strains in their millions. Lower invertebrates can also be cloned — cut an earthworm or flatworm in half, for example, and the missing halves will regenerate to create two genetically identical individuals. Although vertebrates cannot be cloned by these routes, identical twins are naturally occurring genetic clones. Moreover, the method of nuclear transplantation, first developed about 40 years ago in frogs, has been successfully used to make clones of sheep, mice, cows and goats, and it could probably be applied to people too. By taking a few non-reproductive cells from adult mammals, identical replicas can be created without damage (or even inconvenience) to the donors.

But what is the way forward for cloning techniques in the next century? Combined with other cell-biological procedures, they could open the way towards a type of tissue-replacement therapy that avoids the problems of immune rejection. What can we achieve using this methodology, and what are the potential benefits for humans?

How is it done?

In vertebrates, fertilization begins with the union of the sperm and the egg. The unfertilized egg is stopped at a certain stage of the cell-division cycle, and the sperm provides an activation stimulus that triggers the resumption and completion of cell division. The egg and sperm 'pronuclei' then swell, their chromosomes unravel from the tightly packed, 'condensed' state in which they are stored, and DNA replication can proceed. The chromosomes then recondense, the nuclear membrane dissolves, and the fertilized egg divides into two identical daughter cells.

Nuclear transfer subverts fertilization by replacing the female genetic material of an unfertilized egg with the nucleus from a different cell. This was first done successfully on frogs in the 1950s, in the United States and Britain. A non-reproductive (somatic) cell, such as an intestinal epithelial cell, was ruptured by suction into a glass microneedle. Its nucleus, surrounded by a layer of cytoplasm, was then injected into an unfertilized egg from which the female genetic material had

been removed or destroyed by ultraviolet irradiation. Some of these nuclear-transplant embryos developed normally to swimming tadpole or adult stages, and genetic markers were used to show that only the

transplanted nucleus — and not the egg pronucleus — had contributed the genetic material of the resulting embryo¹.

Mechanical disruption of donor cells was also tried in mammals. Until recently, however, it was an unsuccessful or unreproducible method of nuclear transfer. Consistent success in producing live, full-term animals was achieved in livestock by fusing a donor cell from an early embryo with either unfertilized or mock-fertilized eggs as recipients² (Fig. 1). The female genetic material is sucked out, then the recipient cell is fused with the donor nucleus using electrical or chemical methods. In most experiments, the donor nucleus comes from a dividing cell, and it should be able to divide in harmony with the recipient. Nuclear transfer in livestock, using somatic cells as donors and unfertilized eggs as recipients, has been successful (the most famous example being Dolly the sheep), but only in a small percentage of cases³. Live mouse births were initially achieved only by using donor nuclei from very early embryos, although a piezo-electrically controlled microinjection device has facilitated success with the nuclei of adult cells⁴.

How can we tell that mammals obtained by nuclear transfer are genetically identical to the donor cell? Confirmation has come from microsatellite⁵ and DNA fingerprinting⁶ analyses. However, only genetic material in the nucleus will be identical to the nuclear donor. A cell's energy-producing factories, the mitochondria, also contain their own genome. Because mitochondria are located in cytoplasm, they are inherited through the cytoplasm-packed recipient (the egg) rather than through the donor⁷. The mitochondrial DNA in mammalian clones is therefore entirely maternal in origin.

Conservation of the genome

For successful cloning, it is probably essential for donor cells to contain a full complement of genes. Germline cells (the eggs and sperm) have a complete set of genes; that is, those needed for all cell types such as skin cells, intestine and so on. But for nuclear transfer to work, an adult cell that has already been programmed as, say, a skin cell, needs to be somehow reprogrammed so that it regains genetic totipotency — the ability to guide the formation of all the different cell types that make up an animal.

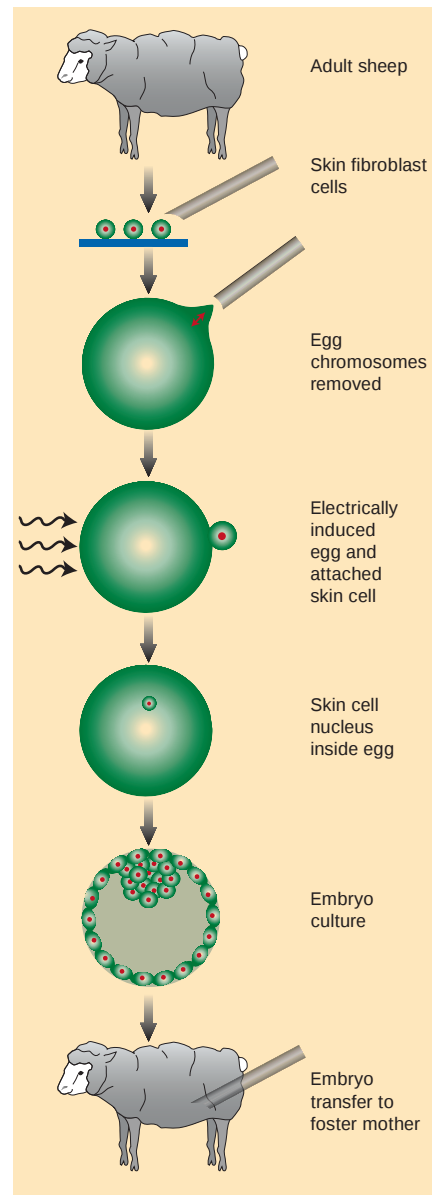


Figure 1 Nuclear-transplantation techniques in mammals. The genetic material is removed from the recipient cell (an egg), then replaced by a nucleus from a donor cell. The resulting embryo is then transferred to a surrogate mother. The clones are genetically identical to the donor.

Early experiments with the frog *Rana pipiens* showed that nuclear totipotency is lost very early in development, at a stage known as gastrulation⁸. However, in another species of frog, *Xenopus laevis*, a transplanted nucleus has been shown to retain genetic totipotency. Indeed, the first clones of adult males and females were obtained in this species from embryonic nuclei (Fig. 2). Although the proportion of successful nuclear transfers decreased as increasingly differentiated donor cells were used (that is, cells more specialized for a particular function), at least a few normal tadpoles (1.5% of the total nuclei transplanted) were obtained from the most advanced donor cells⁹. These experiments culminated in the derivation of genetically marked, fertile male and female adult frogs by transplanting nuclei from intestinal epithelial cells of feeding tadpoles¹⁰. Later experiments used other kinds of differentiated cells as donors, including striated muscle cells, adult skin cells¹¹ and blood cells¹². Thus the principle of totipotency or multipotency was established.

For many years, attempts to transplant nuclei in mammals were unsuccessful, probably because the recipients were fertilized eggs. Nuclear transfer to unfertilized eggs, as in amphibians, has been remarkably successful. Dolly the sheep³ was obtained using cultured mammary-gland cells as donors, and similar experiments with adult mice and cows have yielded fertile adults^{4,13}. Therefore, totipotency of some adult somatic cells has been demonstrated in mammals.

Nevertheless, the yield of normal animals from nuclear transplantation is low. Starting with cells from adult animals, or with fully differentiated cells from tadpoles, only 0.1–1.0% of all eggs receiving transplanted nuclei are born alive (mammals) or reach the swimming stage (tadpoles). Most nuclear transplants either fail to divide or they develop abnormally, possibly due to problems in reprogramming the transplanted nucleus. In amphibians there is the additional problem that incompletely replicated chromosomes are forced to divide prematurely, becoming broken or lost. In mammals, the importance of synchronizing division of the donor nucleus and recipient cell has only recently been appreciated¹⁴.

An important conclusion to come from nuclear-transfer experiments is that the processes of cell differentiation and ageing do not lead to genetic changes in somatic cells. However, there are a few special exceptions. For example, immunoglobulin genes undergo rearrangements to generate antibody diversity. Moreover, mutations can occur in non-reproductive cells, although it is unlikely that changes affecting both copies of an essential gene occur in more than one in 10⁴ cells. Finally, telomeres, the specialized structures at the ends of chromosomes, shorten progressively with age in somatic



Figure 2 Cloned frogs. These 19 identical male albino frogs were prepared by nuclear transplantation into unfertilized eggs of the dark green female frog³³. (Male frogs are about half the size of females.)

cells. In mice unable to generate telomeres, however, it takes several generations before this has any noticeable effect¹⁵. Moreover, sheep made through somatic nuclear transfer have shorter telomeres, but this has had no detectable effects so far⁷.

Gene reprogramming

The pattern of gene expression in adult cells is very different from that in embryonic cells. In amphibians, for example, a number of genes expressed in embryos five hours after fertilization are not expressed in differentiated (specialized) larval or adult cells¹⁶. Conversely, some genes are expressed in adult cells but not in early embryos. When embryos are analysed a few hours after the transfer of adult cell nuclei, gene expression cannot be distinguished from that in embryos grown from fertilized eggs¹⁷. This means that the exchange of cytoplasm around a nucleus, from that of an adult cell to that of an egg, causes a dramatic switch in gene expression in only a few hours. A nucleus that was once part of an intestine, skin or muscle cell is therefore transformed into that of an embryonic cell.

The situation is probably different for imprinted genes in mammals. Imprinting, which is required for normal development, is a process by which about 30 genes are marked during sperm and egg formation such that they are switched off in the embryo¹⁸. Imprinted genes are unlikely to be reprogrammed by nuclear transfer, because embryos would not survive if they were. Moreover, imprinting is normally reversed during egg and sperm formation, rather than after fertilization.

Another focus of reprogramming attention is the inactive X chromosome. During early development of female mammals, one of the two X chromosomes is randomly inactivated in those tissues contributing to the fetus¹⁹. However, in the embryonic tissues that contribute to the placenta, the paternal X chromosome is always the one that is inactivated¹⁹. These observations

prompt a question, so far unanswered, concerning the use of female cells in cloning. Will the inactive X chromosome become reactivated in the tissues of the newly created animal? If not, will embryos reconstructed from a donor containing an inactive maternal X chromosome be viable, given that the maternal X chromosome is usually inactivated in the paternal tissue (which will therefore have no active X-chromosomal genes)?

In amphibians, reprogramming of gene expression is accompanied by massive enlargement of the nucleus and exchange of proteins with those in the cytoplasm. When a nucleus is transplanted to a frog egg, it undergoes 12 rounds of division before new gene activity commences. Does gene reprogramming require the formation of new DNA (DNA replication)? To test this, several nuclei can be injected at once into growing egg cells in which replication does not take place. The transplanted nuclei nevertheless change their pattern of gene expression to conform with that of a growing egg cell²⁰. These experiments show that gene reprogramming can occur on the same actual genes as were present before nuclear transfer, and does not require the formation of new DNA. Key molecules found in eggs that may bring about reprogramming include nucleoplasmin and certain embryo-specific histones (proteins around which the DNA is wrapped).

New tissues for old

Damaged or diseased tissues often cannot be repaired by drugs or other medication, and most organs and tissues regenerate very poorly in mammals. In a few cases, artificial materials such as replacement joints, or mechanical devices such as renal dialysis machines, work remarkably well. But ultimately, the most satisfactory remedy is to transplant organs or tissues from other people.

Transplantation works reasonably well for kidneys, hearts and so on, but there are three major disadvantages. First, the supply of such organs is extremely limited, depending largely on donations from accident victims. Second, treatment is very expensive (about £100,000, or US\$160,000, for a replacement heart). Third, recipients need to be given immunosuppressive drugs to avoid rejection of the transplanted organ owing to the genetic differences between donor and recipient. Although the use of animal organs has been considered for transplantation, the genetic incompatibility is even greater with these, and even animal organs that have been engineered to contain human immune regulator genes are still targets for rapid rejection.

An alternative strategy involves stem cells; these are cells that can renew themselves and also give rise to a variety of differentiated cell types. Mammalian bone

marrow, for example, contains a range of haematopoietic (blood-forming) stem cells. Some of these can be isolated and encouraged to proliferate using natural signalling molecules such as members of the interleukin family. These stem cells can then be made to progress to form more restricted stem cells (which can form a more limited range of cell types) and, eventually, to form fully differentiated cells such as erythrocytes or granulocytes. This type of blood stem-cell therapy has been practised for many years in humans, but the quantity and quality of available stem-cell types is very limited. The prospects also seem good for stem cells obtained from olfactory placodes. These neural stem cells can be proliferated in culture, and they have been shown to restore function in the mouse central nervous system.

It is also possible for stem cells of one type to generate occasional cells of a different type, although the conditions that achieve this are not known. For example, neural stem cells can generate haematopoietic stem cells when transplanted to mice that have been irradiated to eliminate their own blood stem cells²¹. Stem cells from human bone marrow have been reported to generate functional neural cells²². An ideal stem cell is that exemplified by mouse embryonic stem or germ cells, which are derived experimentally from early mouse embryos or germline cells respectively. These cells can be proliferated in culture and, when transplanted to hosts, can differentiate into all types of adult cell. Human cells with several properties of mouse embryonic stem cells have also been described^{23,24}. However, all embryonic stem and germ cells are currently obtained by killing normally generated early embryos, raising ethical concerns for human material.

Despite their advantages, transplantation and stem-cell strategies still suffer from the problem of immunological incompatibility, a problem that could be avoided if material derived from a patient's own tissues could be transplanted into them. This is already done with skin for burns patients, but the amount of material is very limited and, for most tissues, it is not yet possible to obtain enough stem cells, if they can be obtained at all. One approach to the problem is to freeze, at birth, samples of cells from the umbilical cord. This tissue is rich in stem cells which, after proliferation and differentiation, might be of use later in life.

Therapeutic cloning

Given the problems of rejection, the lack of identified stem cells for most tissues, and the difficulties of using normal human embryos as a source of embryonic stem cells, an altogether different route is to combine therapeutic cloning with the use of differentiation factors (Fig. 3). Although not yet practicable,

this scheme offers a realistic possibility for the future.

The first stage is to use nuclear transplantation to reprogramme the nucleus of a human adult cell and obtain a blastocyst (a very early embryo). This step, in effect, transforms an adult cell into an embryonic cell of the same genetic constitution. By doing a number of such transplants in series, the yield of embryos will be increased. The next

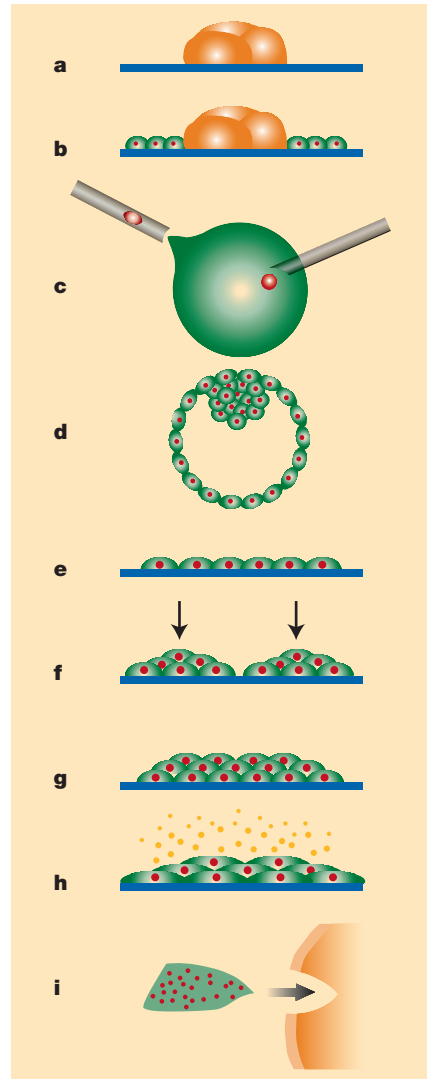


Figure 3 Steps involved in therapeutic cloning.

a, An explant is taken from an adult tissue donor. b, The tissue sample is grown up in culture (cell proliferation). c, Nuclei from the outgrowth cells are transplanted into enucleated eggs (eggs from which the genetic material has been removed). d, The nuclear-transplant eggs are cultured to the blastocyst stage. e, Embryonic stem-cell lines are derived from the inner cell mass. f, The cells are transfected with genes encoding proteins that will cause cell suicide if cells begin to proliferate out of control. g, Selected cells begin to proliferate. h, Differentiation factors are added and the cells begin to differentiate into specialized cell types. i, The replacement tissue is transplanted into the original donor.

stage is to expand this population of embryonic cells in culture by promoting proliferation but not differentiation. This has not yet been done with embryos obtained by nuclear transplantation, but it works well for mouse and human embryonic stem-cell-like cells, showing that it can be done in principle.

Having obtained a certain number of cells, the next step is to differentiate them in a desired direction. In amphibians this works well when blastula cells are exposed to signalling molecules that act early in development, including bone-morphogenetic proteins, the fibroblast growth factor, Cerberus and noggin. A molecule called activin, which belongs to the transforming growth factor- β family, is particularly useful. Activin is very stable, easy to make in a biologically active form, effective at extremely low concentrations, and it can direct blastula cells towards a wide variety of cell types depending on the concentration at which it is used²⁵. Combinations of other secreted factors generate an even greater variety of differentiated cell types. A mass of differentiated cells could then be transplanted back to the original nuclear donor.

Might such extensive manipulations cause some of the resulting cells to become cancerous? Possibly, but this could, in principle, be avoided, at least for those tissues that are not continuously replaced in the body. If a gene (or genes) were introduced to limit the number of divisions that the cells may undergo before their terminal differentiation (a state in which they can no longer divide), cells would be unable to proliferate out of control. Another way to achieve this objective could be to reduce the length of telomeres during *in vitro* culture.

Of course, several steps will need further development before this procedure is of practical use. These include:

- An increased efficiency of nuclear transfer.
- The ability to derive a population of embryonic stem-cell-like cells from nuclear-transplant blastocysts.
- An increased ability to differentiate such embryonic stem cells into functional tissues.
- Methods to screen for any cell lines that may have incurred genetic damage, and that would not therefore be suitable for human *in vivo* use.

This scheme has several advantages. It requires unfertilized eggs rather than embryos, and the nuclear-transplant blastocysts do not need to be able to develop normally (they only need to form the required cell types). Moreover, populations of cells or simple tissues may be useful, avoiding the need to create complex organs. Above all, the newly generated cells would be almost totally compatible with the recipient. We anticipate that any incompatibilities, owing to minor histocompatibility antigens in the transplanted cells, which will differ from those of



Figure 4 Countries in which anti-cloning legislation has been, or will shortly be, passed. Legislation has been passed in countries shaded green, and is pending in countries shaded yellow. Countries with red stripes are co-signatories to the January 1998 Council of Europe Protocol. This must be ratified by at least five countries. Some (although not all) of this legislation applies equally to both reproductive and therapeutic cloning.

the patient, should not be significant. This supposition could easily be tested by skin transplantation between a group of cloned animals¹³.

Legality and ethics

The publication of Dolly the sheep's existence had an immediate and generally negative effect on public opinion worldwide, and sparked off a flurry of restrictive legislative activity. Most invective was aimed at the prospect of human reproductive cloning; that is, the production of a person using nuclear transfer. Both safety and ethical reasons were cited against such cloning. Belated sober reflection revealed that few of these moral arguments were sound. For example, the ethical objection about the welfare of cloned children has been shown to apply equally well to the case of natural procreation²⁶: if the anticipated psychological damage to clones from societal prejudice is enough to ban the practice, then the psychological damage to a mixed-race child from racism should justify a like ban on procreation from mixed-race couples. Clearly the latter would be absurd.

Therapeutic cloning is ethically less contentious because a new person is not produced. However, as for abortion, the issue of the deliberate destruction of a potential person is raised²⁷. Therapeutic cloning is illegal in many countries (Fig. 4), where legislation prevents the use of human eggs for therapeutic research. This prohibits therapeutic cloning using unfertilized eggs, even though a nuclear-transplant embryo would be used to make embryonic stem-cell-like cells, and so would have no potential for survival because it would not be implanted.

No such legislation has yet been enacted in the United States, except that federal funds cannot be used for reproductive cloning, nor

can they be used in generating embryos for therapeutic cloning. They can, however, be used for research on cell lines derived from human embryos. In the United Kingdom it is legal to use human embryos up to 14 days old for research connected with fertility or contraception. It is illegal to use human eggs for any purpose where the intention is to create an embryo, even if only for cell replacement. Permission to do this (and, hence, to allow therapeutic cloning) could be granted by the Secretary of State for Science under an addition to the present Human Embryology Act.

Perhaps a way round these issues is to use cells or eggs from other species. However, as yet, normal development has not been obtained with nuclear transfers between species. All but one of the human-nucleus to cow-egg transfers attempted died before the embryo had divided five times, and other mammalian combinations behaved similarly^{28,29}. It is therefore unlikely that animal eggs could be used as an alternative to human eggs as recipients for human nuclei.

The future of cloning

Live births have been achieved using somatic nuclear transfer in mice, sheep, cows and goats, and the technique would probably also be successful in humans. However, the low efficiency of the procedure may restrict its immediate application to the production of small herds of identical animals. There will be special applications, though, where the production of single animals will be useful. For example, the technique could be used to inactivate prion genes in ruminants³⁰ (scrapie/BSE research), to disrupt the $\alpha 1$ -3 galactosyl transferase gene in pigs³¹ (for xenotransplantation), and to knock out the cystic-fibrosis transmembrane-conductance-regulator gene in sheep³² (creating an animal 'model' in which this disease could be

studied). Nuclear transfer is currently the only option for gene targeting in livestock, and one of us (A.C.) has shown that sheep can be generated with targeted changes in particular genes.

Our belief, however, is that the greatest eventual benefit of the new technology will be in therapeutic cloning; the use of somatic-cell nuclear transfer to generate replacement tissues or organs. This would avoid the risks of tissue rejection by supplying a person with new tissue of exactly their own genetic type. All of the main steps in the therapeutic cloning procedure have been achieved individually, albeit at a low efficiency. We now need to improve the success rate for generating nuclear-transplant embryos from adult tissues, to find ways of generating embryonic stem-cell cultures from nuclear-transplant embryos, and to control more accurately the pathways of stem-cell differentiation and the formation of whole organs *in vitro*. As soon as any new scientific technique works at all, it is almost always improved in both efficiency and ease of operation. This seems likely to be the case for cloning technology too.

J. B. Gurdon is at the Wellcome/CRC Institute, Tennis Court Road, and in the Department of Zoology, University of Cambridge, Cambridge CB2 1QR, UK.
e-mail: j.b.gurdon@welc.cam.ac.uk

Alan Colman is at PPL Therapeutics, Roslin Institute, Edinburgh EH25 9PP, UK.
e-mail: acolman@ppl-therapeutics.com

1. Elsdale, T. R., Gurdon, J. B. & Fischberg, M. *J. Embryol. Exp. Morph.* **8**, 437–444 (1960).
2. Sun, F. Z. & Moor, R. *Curr. Top. Dev. Biol.* **30**, 145–173 (1995).
3. Wilmut, I. *et al. Nature* **385**, 810–813 (1997); erratum, *Nature* **386**, 200 (1997).
4. Wakayama, T. *et al. Nature* **394**, 369–374 (1998).
5. Ashworth, D. *et al. Nature* **394**, 329 (1998).
6. Signer, E. *et al. Nature* **394**, 329–330 (1995).
7. Shiels, P. G. *et al. Nature* **399**, 316–317 (1999).
8. Briggs, R. & King, T. J. *J. Morphol.* **100**, 269–312 (1957).
9. Gurdon, J. B. *J. Embryol. Exp. Morphol.* **10**, 622–640 (1962).
10. Gurdon, J. B. & Uehlinger, V. *Nature* **210**, 1240–1241 (1966).
11. Gurdon, J. B., Laskey, R. A. & Reeves, O. R. *J. Embryol. Exp. Morphol.* **34**, 93–112 (1975).
12. DiBerardino, M. A. & Hoffner, N. J. *Science* **219**, 862–864 (1983).
13. Wells, D. N., Missa, P. M. & Tervit, H. R. *Biol. Reprod.* **60**, 996–1005 (1999).
14. Campbell, K. H., Ritchie, W. A. & Wilmut, I. *Biol. Reprod.* **49**, 933–942 (1993).
15. Blasco, M. A. *et al. Cell* **91**, 24–34 (1997).
16. Stennard, F., Ryan, K. & Gurdon, J. B. *Curr. Opin. Genet. Dev.* **7**, 620–627 (1997).
17. Chan, A. P. & Gurdon, J. B. *Int. J. Dev. Biol.* **40**, 441–451 (1996).
18. Surani, M. A. *Cell* **93**, 309–312 (1998).
19. Heard, E., Clerc, P. & Auner, P. *Annu. Rev. Genet.* **31**, 571–610 (1997).
20. De Robertis, E. M., Partington, G. A., Longthorne, R. F. & Gurdon, J. B. *J. Embryol. Exp. Morphol.* **40**, 199–214 (1977).
21. Bjornson, C. *et al. Science* **283**, 534–537 (1999).
22. Azizi, S. A. *et al. Proc. Natl Acad. Sci. USA* **95**, 3908–3913 (1999).
23. Thomson, J. A. *et al. Science* **282**, 1145–1147 (1998).
24. Shambloott, M. J. *et al. Proc. Natl Acad. Sci. USA* **95**, 13726–13731 (1998).
25. Green, J. B. *et al. Cell* **71**, 731–739 (1992).
26. Burley, J. & Harris, J. *J. Med. Ethics* **25**, 108–113 (1999).
27. Burley, J. *Semin. Cell Dev. Biol.* **10**, 287–294 (1999).
28. Lanza, R. P., Cibelli, J. B. & West, M. D. *Nature Med.* **5**, 975–977 (1999).
29. Dominko, T. *et al. Biol. Reprod.* **60**, 1496–1502 (1999).
30. Bueler, H. *et al. Cell* **73**, 1339–1347 (1993).
31. Tearle, R. G. *et al. Transplantation* **61**, 13–19 (1996).
32. Snouwaert, J. N. *et al. Science* **257**, 1083–1088 (1992).
33. Gurdon, J. B. *Proc. R. Soc. Lond. B Biol.* **198**, 211–247 (1977).

Large eyeballs in diving ichthyosaurs

The huge eyes of these extinct reptiles may have been useful deep in the ocean.

Ichthyosaurs are marine reptiles that existed about 250 million to 90 million years ago. They had fish-shaped bodies, which were exceptional among Mesozoic marine reptiles¹. Here we report that ichthyosaurs also had huge eyeballs — larger than those of any other vertebrate. We infer that the genus *Ophthalmosaurus*, whose eyes were particularly large and sensitive, used to dive to depths of 500 metres or more.

Absolute size is an important property of eyes^{2–4}, because larger eyes can house more retinal photoreceptive cells and receive more light per solid angle of image space^{2–4}. Eye size also usually reflects the importance of vision in animals⁴: for example, the horse has among the largest eyeballs of any land animal alive today⁴, about 50 mm across, which may be important, given its fast speed³. But eye size also scales with body size³ — for example, the blue whale has the largest eyes of any living vertebrate⁴, about 150 mm across, although this is small for such a colossal body. These scaling effects should be considered when discussing eye size.

We used the sclerotic ring diameter to estimate the eyeball diameter of parvipelvians, an ichthyosaur group with tuna-shaped bodies⁵ (Fig. 1), and compared its scaling with other tetrapods (Fig. 2). Eyeball diameters of tetrapods of a given body size are usually restricted within a narrow range³ (Fig. 2). Parvipelvian ichthyosaurs, and some birds with sensitive vision, did not share this constraint, having large eyes relative to body length (Fig. 2).

The largest sclerotic ring we examined, 253 mm in its external diameter, belongs to *Temnodontosaurus*, which had a body length of about 9 m. There also is a poorly known parvipelvian ichthyosaur that may have been 15 m long⁶, so the largest ichthyosaurian eye was probably more than 300 mm in diameter. The giant squid *Architeuthis* is thought to have the largest eyeball of any extant animal, having been estimated as approaching 250 mm in diameter⁷.

Ophthalmosaurus had the largest eyes (more than 220 mm in diameter) of any ichthyosaur for its body length (Fig. 2), and the largest sclerotic ring aperture, with a diameter of about 100 mm. We estimated the



Figure 1 Artistic impression of the ichthyosaur *Ophthalmosaurus*.

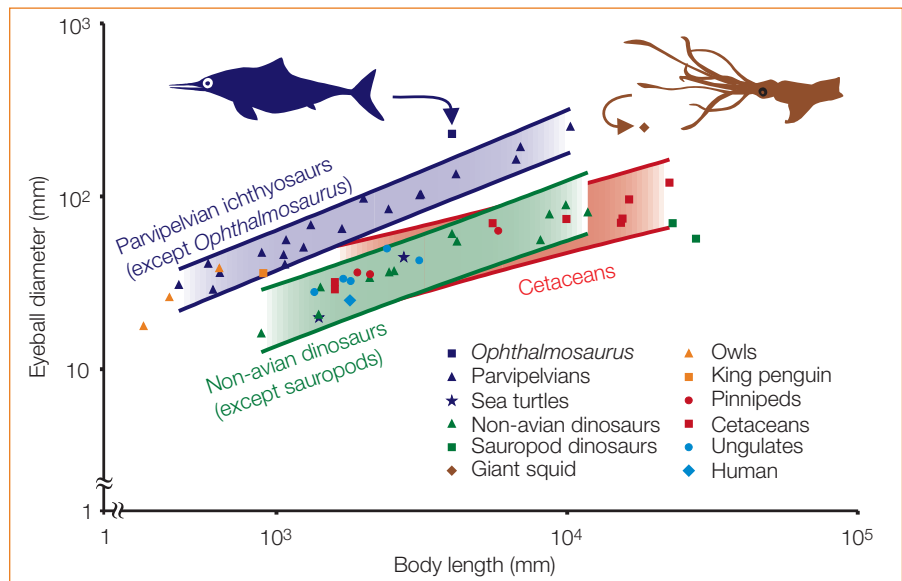


Figure 2 Logarithmic plot of eyeball diameter against body length. Bands show 95% confidence ranges for the least-square regression lines for parvipelvian ichthyosaurs except *Ophthalmosaurus* ($n=19$), non-avian dinosaurs except sauropods ($n=12$), and cetaceans ($n=8$). Eyeball diameters for dinosaurs and ichthyosaurs are based on the external diameter of the sclerotic rings (which were sometimes estimated from the height of the orbit in ichthyosaurs). Data for non-ichthyosaurs were derived from the literature.

minimum f -number (the same measure of relative aperture as is used for camera lenses) of several ichthyosaur eyes (see Supplementary Information). The minimum f -numbers for typically nocturnal and diurnal vertebrate eyes are about 0.95 and 2.1, respectively³. The minimum f -number of an *Ophthalmosaurus* eye, the lowest of any ichthyosaur, was calculated to be between 0.76 and 1.1, so the genus seems to have been capable of seeing in low-light conditions.

The large sclerotic ring aperture indicates that *Ophthalmosaurus* could probably detect point light sources, such as luminance from the photophores of prey⁸, a useful ability in the mesopelagic layer of the ocean (depths of 200 to 1,000 m). Cats, whose eyes have a similar minimum f -number, could theoretically see to a depth of 500 m in most oceans (based on data from refs 8–10; see Supplementary Information). *Ophthalmosaurus* was roughly 4 m long, with a mass of 930 kg, comparable to the size of living mesopelagic diving animals¹¹. Conservative estimates of diving duration and swimming speed indicate that it could dive to a depth of 600 m (see Supplementary Information).

To test whether *Ophthalmosaurus* was a deep diver, we examined the frequency of pathology arising from Caisson disease (also known as the 'bends') in the humeri and femora of various ichthyosaurs (see Supplementary Information). Deep-diving animals do not usually suffer from bends¹²,

although some turtles do when an accident or escape response forces them to depart from their normal diving pattern, causing a high partial pressure of carbon dioxide in the blood¹³. We found that the two genera with the lowest minimum f -numbers had the highest frequencies of the bends.

Ryosuke Motani*, **Bruce M. Rothschild†**, **William Wahl Jr‡**

*University of California Museum of Paleontology, Berkeley, California 94720, USA

†The Arthritis Center of Northeast Ohio, Youngstown, Ohio 44512, USA

‡Tate Geological Museum, Casper, Wyoming 82601, USA

§Present address: Department of Paleobiology, Royal Ontario Museum, Toronto, Ontario M5S 2C6, Canada
e-mail: ryo.motani@utoronto.ca

1. Motani, R., You, H. & McGowan, C. *Nature* **382**, 347–348 (1996).
2. Martin, G. R. *Prog. Sens. Physiol.* **4**, 43–81 (1983).
3. Hughes, A. in *The Visual System in Vertebrates* (ed. Crescitelli, F.) 613–756 (Springer, Berlin, 1977).
4. Walls, G. L. *The Vertebrate Eye and its Adaptive Radiation* (Cranbrook, Bloomfield Hills, 1942).
5. Motani, R. *J. Vert. Paleontol.* **19**, 472–495 (1999).
6. McGowan, C. *Can. J. Earth Sci.* **33**, 1011–1021 (1996).
7. Roper, C. F. E. & Boss, K. J. *Sci. Am.* **246**, 96–105 (1982).
8. Martin, G. R. *Ibis* **141**, 444–450 (1999).
9. Jerlov, N. G. *Marine Optics* (Elsevier, Amsterdam, 1976).
10. Gunter, R. J. *Physiol. (Lond.)* **114**, 8–15 (1951).
11. Schreer, J. F. & Kovacs, K. M. *Can. J. Zool.* **75**, 339–358 (1997).
12. Kooyman, G. L. & Ponganis, P. J. *Annu. Rev. Physiol.* **60**, 19–32 (1998).
13. Rothschild, B. M. *Comp. Biochem. Physiol.* **100**, 529–535 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Developmental biology

BMP signalling specifies the pyloric sphincter

Sphincters are muscular valves that form at the boundaries between organs of the gut; for example, the pyloric sphincter forms at the junction of the small intestine and stomach. We show here that signalling by bone morphogenetic protein (BMP) from the avian small intestine induces the cells of the adjacent gizzard (the equivalent of the stomach in the chicken) primordium to form a sphincter. This finding and related studies of the role of Hox genes in the specification of the ileocaecal sphincter¹ provide insights into the processes by which new cell fates are specified at the borders between distinct embryological domains.

To investigate the inductive signalling involved in the development of the chicken gut, we examined the expression pattern of several members of the BMP family, including BMP-2, BMP-4, BMP-5 and BMP-7, and compared them with the expression domains of the known BMP receptors². BMP-4 is the only member of the family to be expressed in the early chicken gut, appearing in the mesoderm of the small intestine from embryonic day 2.5 (E2.5) (Fig. 1a,b, and data not shown).

Several BMP receptors are expressed in the intestinal mesoderm in a position to mediate BMP-4 signalling in this portion of the gut (data not shown). However, the type I receptor, *BMPRI1B*^{3,4}, is specifically expressed in the gizzard mesoderm from E2.5 (Fig. 1c), despite the fact that neither BMP-4 nor any other related BMP family member we examined (data not shown) are expressed in the early gizzard mesoderm.

One possible explanation is that the expression of this receptor in the developing gizzard reflects a role in transducing BMP signals originating from adjacent regions of the gut, such as the developing small intestine where BMP-4 is expressed, thereby mediating the patterning of structures at the border of the gizzard and the small intestine.

The pyloric sphincter forms at the junction of the gizzard and the small intestine, and acts as a gate to regulate the passage of ingested food from the stomach to the small intestine. The sphincter differs from the rest of the gizzard with regard to the mesodermal and endodermal layers⁵.

While searching for a molecular marker for the pyloric sphincter, we examined the expression pattern of *Nkx2.5*, a homeo-domain-containing transcription factor that is expressed in the midgut⁶. We found that *Nkx2.5* is a specific marker for the mesoderm of the pyloric sphincter in the chick embryo (Fig. 1e,f). In the early stages, it is expressed adjacent to the area where BMP-4 is expressed, and overlaps with the expression pattern of *BMPRI1B* in the posterior of its domain (Fig. 1b–d).

To investigate the role of BMP-4 in the formation of the pyloric sphincter, we injected a retrovirus containing the *mBMP-4* complementary DNA⁷ into stage-10 chick embryos⁸. Ectopic expression of *Nkx2.5* was seen in the gizzard, but not in the small intestine, of injected embryos, even though both organs were infected with the retrovirus (Fig. 1i,j). Sections through infected gizzards showed that *Nkx2.5*, like the endogenous domain, is limited to the mesodermal layer (Fig. 1g,h). BMP signalling is therefore sufficient to activate *Nkx2.5* in the gizzard primordium.

We tested whether BMP signalling is necessary for endogenous *Nkx2.5* activation in the developing sphincter by injecting a virus containing the secreted BMP antagonist *Noggin*⁹ into chick embryos ($n=12$). *Nkx2.5* was downregulated at the border of the gizzard and the small intestine in injected embryos (Fig. 1k), indicating that BMP signalling is required for *Nkx2.5* expression in the gut. *Noggin* is a specific antagonist of BMP-2, -4 and -7 (ref. 10), of which only BMP-4 is expressed in this region of the gut, indicating that the endogenous signal for *Nkx2.5* induction is BMP-4 from the small intestine. *Noggin* misexpression also altered development in other regions of the gut (data not shown), reflecting different roles of BMP signalling in gut patterning².

To test the morphological consequence of altering BMP signalling during the establishment of the pyloric sphincter, we allowed BMP4-infected embryos to develop to E9 (*Noggin*-infected embryos did not survive long enough to allow morphological analysis of guts where BMP signalling was disrupted). In addition to its role during sphincter specification, prolonged BMP expression interferes with muscle development and affects the rates of mesodermal proliferation and apoptosis² (data not shown). We therefore limited our morphological analysis to the endoderm.

Although the gizzard endoderm has long, thin villi that are completely covered by a thick layer of a protective keratin-like substance⁷ (Fig. 1l), the pyloric sphincter endoderm has short villi, each of which is thin at the base and bulbous at the tip (Fig. 1m). Instead of the long, keratin-covered villi characteristic of the gizzard, the endoderm of BMP4-infected gizzards had short villi with thin bases and bulbous extensions,

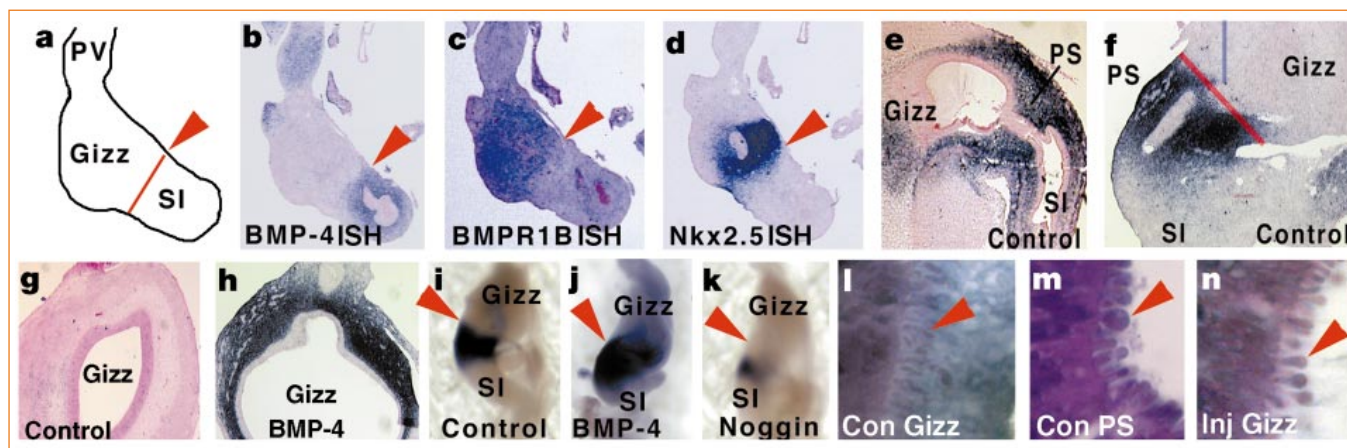


Figure 1 BMP signalling is necessary and sufficient to specify a pyloric sphincter. **a**, Diagram of the regions of the embryonic gut. **b–d**, Section *in situ* hybridization (ISH)² of an E4.5 gut with adjacent sections stained by using riboprobes for BMP-4 (**b**), *BMPRI1B* (**c**) or *Nkx2.5* (**d**). Red arrowheads point to the posterior border of expression of *Nkx2.5* and *BMPRI1B* as well as the anterior border of BMP-4 expression. **e**, Section ISH of an E9.5 chicken gut using an *Nkx2.5* riboprobe. **f–h**, Section ISH of E6 gizzard sections hybridized with an *Nkx2.5* riboprobe on a control gut (**f**, **g**), or an embryo injected with RCAS-BMP-4 (ref. 7) (**h**) showing ectopic *Nkx2.5* expression in the gizzard. Red line in **f** demarcates the anterior boundary of *Nkx2.5* expression in wild-type embryos; in **g** and **f**, the blue line shows the plane of sectioning. **i–k**, Whole-mount ISH¹¹ using *Nkx2.5* as a riboprobe on an E4.5 control gut (**i**), an E4.5 embryo injected with RCAS-BMP-4 showing ectopic *Nkx2.5* in the gizzard (**j**), or an E4.5 *Noggin*⁹-injected embryo showing a decrease in *Nkx2.5* expression at the pyloric sphincter (**k**). The red arrowheads indicate the anterior border of *Nkx2.5* expression in a control gut. **l–n**, Sections through an E9 embryo showing morphology of the endoderm of either the gizzard (**l**), pyloric sphincter (**m**) or the gizzard (**n**) of an embryo injected with BMP-4 with villi resembling that of the pyloric sphincter. Red arrowheads point to villi at the tips of endodermal cells. Gizz, gizzard; SI, small intestine; ISH, *in situ* hybridization; PS, pyloric sphincter; PV, proventricularis; inj, injected; con, control.

devoid of overlying keratin (Fig. 1n), resembling those of the pyloric sphincter.

Our results indicate that BMP-4 from the intestinal mesoderm acts on the adjacent region of the gizzard to induce the expression of *Nkx2.5* and to regulate the formation of the pyloric sphincter. BMP signalling is therefore important in the regionalization of the gut, specifying the sphincter at its proper location to act as a gate between the gizzard and intestine.

Devyn M. Smith, Clifford J. Tabin

Department of Genetics, Harvard Medical School,
200 Longwood Avenue, Boston,
Massachusetts 02115, USA
e-mail: tabin@rascal.med.harvard.edu

1. Zakany, J. & Duboule, D. *Nature* **401**, 761–762 (1999).
2. Roberts, D. J., Smith, D. M., Goff, D. & Tabin, C. J. *Development* **125**, 2791–2801 (1998).
3. Zou, H., Wieser, R., Massagué, J. & Niswander, L. *Genes Dev.* **11**, 2191–2203 (1997).
4. Natsumi, T. et al. *J. Biol. Chem.* **272**, 11535–11540 (1997).
5. Romanoff, A. L. *The Avian Embryo* (Macmillan, New York, 1960).
6. Buchberger, A. et al. *Mech. Dev.* **56**, 151–163 (1996).
7. Duprez, D. et al. *Mech. Dev.* **57**, 145–157 (1996).
8. Hamburger, V. & Hamilton, H. L. *J. Morphol.* **88**, 49–92 (1951).
9. Capdevila, J. & Johnson, R. L. *Dev. Biol.* **197**, 205–217 (1997).
10. Zimmerman, L. B., DeJesus-Escobar, J. M. & Harland, R. M. *Cell* **96**, 599–606 (1996).
11. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. *Cell* **75**, 1401–1416 (1993).

Chronometry

Effect of the 1999 solar eclipse on atomic clocks

Solar eclipses have been reported to have a strange influence on the behaviour of atomic clocks¹ and pendulums^{2,3}, which has been attributed to some unknown feature of gravity⁴. Here we correct this idea after being unable to detect any anomalous changes in the relative rates of three types of atomic clock, based on the ground-state hyperfine transitions of hydrogen, rubidium and caesium, during the solar eclipse of 11 August 1999 over central Europe.

The reported effects of solar eclipses on atomic clocks and the movement of pendulums have been mixed. A study of the relative rates of rubidium and caesium atomic clocks during four partial solar eclipses between 1987 and 1992 (ref. 1) recorded an accumulated time difference between two caesium clocks and two rubidium clocks located in the same laboratory of 468 ns and 65 μ s, respectively. In these experiments, the change in the clock rates occurred not only during the eclipse, but also up to one day before and after the maximum of the eclipse.

Peculiar movements of a Foucault pendulum were reported at the time of the onset of solar eclipses in 1954 and 1959 (ref. 2). A variation in the oscillation period of torsion pendulums allegedly occurred

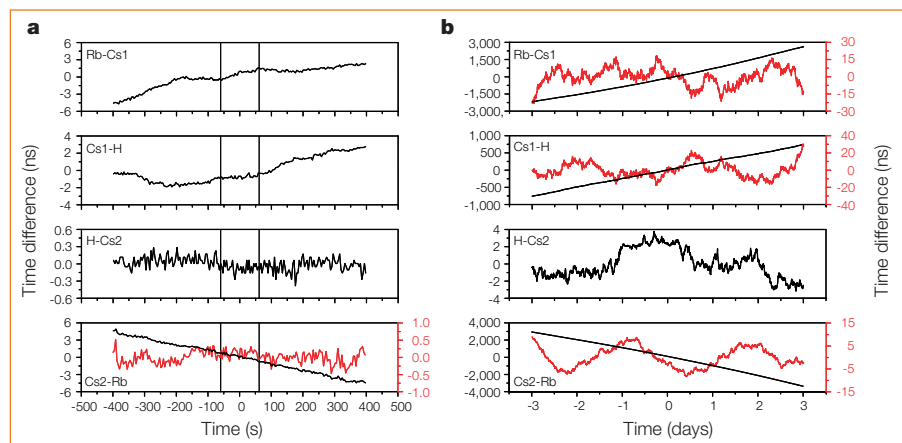


Figure 1 Time differences in four atomic clocks during the solar eclipse. Rb-Cs1, Cs1-H, H-Cs2, Cs2-Rb are the time differences between two caesium clocks, a rubidium clock and a hydrogen maser. **a**, Time period of 800 s around the eclipse. Vertical lines mark the beginning and end of the total eclipse; time 0 indicates the maximum of the eclipse. **b**, Time period of 6 days. Some plots show a smooth drift because the clocks are not perfect. Red curves are obtained by removing this drift by fitting a low-order polynomial.

during the onset of an eclipse³, although this was never confirmed^{5–7}. In addition, there has been a report of a highly significant variation in the gravitational field during the onset of a solar eclipse⁸.

Although general relativity predicts that gravity can influence clock rates, it does not explain why clocks in the same place should be affected differently. It is not clear why this effect should occur when the Sun is eclipsed by the Moon, and not just when the Moon is new and the Sun and Moon are almost aligned, unless gravity is shielded. But then, why does it not occur every day, given that the Sun is shielded by the Earth once every revolution⁹? A change in the value of the fine-structure constant during the eclipse might explain the changes in clock rate.

To try and make sense of these observations, we repeated the experiments of Zhou et al.¹ by comparing four atomic clocks during the 1999 total solar eclipse over central Europe. The clocks, based on ground-state hyperfine splitting of different species of atom, were placed in an air-conditioned room in a basement at Wessling in Germany (48° 05' 03" N, 11° 16' 40" E, 620 m above sea level). Their frequency depends differently, with known coefficients, on the value of the fine-structure constant¹⁰. A computer recorded their time difference every 4 seconds from 3 August to 23 August.

To exclude the influence of any spurious effects, such as variations in the background magnetic field (resulting from changes in the ionosphere, for example), we recorded variations in temperature, power lines and barometric pressure. We detected increased magnetic-field fluctuations of up to ± 0.1 microtesla in the vertical direction during working hours from Monday to Friday as a result of human activity. We recorded the external light intensity to verify synchronization of the data with the eclipse. The time axis was constantly calibrated by a

time signal broadcast by the Physikalisch-Technische Bundesanstalt in Braunschweig and zeroed at 10:37:46 universal time, the maximum of the eclipse. No significant changes in these parameters seemed to be connected with the eclipse.

The relative time differences of four atomic clocks (one rubidium and two caesium clocks and one hydrogen maser) are shown in Fig. 1 for time intervals of 800 seconds and 6 days. Phase drifts are visible in some cases, depending on the type of clock and the duration of the observation (black curves), indicating that the clocks are not perfectly stable in their relative rates. To reveal possible effects of the eclipse, we fitted these drifts with polynomials of the order of less than four to obtain the residue shown in red. To enable comparison with the results of Zhou et al.¹, who observed changed clock rates for 8 hours in one case and 3 days in another, we plotted our data over a period of 6 days.

We can set an upper limit, independent of the type of clock, of ± 6 ns for a non-regular time difference accumulated over 800 seconds, and ± 20 ns accumulated over 6 days. It could be argued that some features might have been suppressed by the way the data were handled, so the complete data set is available at <http://www.mpq.mpg.de/~haensch/eclipse/eclipse.html>

Thomas Udem*, Jörg Reichert*, Ronald Holzwarth*, Theodor Hänsch*, Rainer Krämer†, Jörg Hahn†, Jens Hammesfahr†

*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, 85748 Garching, Germany

†Deutsches Zentrum für Luft- und Raumfahrt e.V., Münchner Strasse 20, 82230 Wessling, Germany

1. Zhou, S. W., Huang, B. J. & Ren, Z. M. *Nuovo Cimento C* **18**, 223–236 (1995).
2. Allais, M. *Aero/Space Eng.* **9**, 46–55 (1959).
3. Saxl, E. J. & Allen, M. *Phys. Rev. D* **3**, 823–825 (1971).
4. Eckhard, D. H. *Phys. Rev. D* **42**, 2144–2145 (1990).

5. Jun, L. *et al.* *Phys. Rev. D* **44**, 2611–2613 (1991).
6. Kuusela, T. *Phys. Rev. D* **43**, 2041–2043 (1991).
7. Kuusela, T. *Gen. Relativ. Grav.* **24**, 543–551 (1992).
8. Mishra, D. C. & Rao, M. B. S. *Curr. Sci.* **72**, 782–783 (1997).
9. Turneaure, J. P., Will, C. M., Farrell, B. F., Mattison, E. M. & Vessot, R. F. C. *Phys. Rev. D* **27**, 1705–1714 (1983).
10. Prestage, J. D., Tjoelker, R. L. & Maleki, L. *Phys. Rev. Lett.* **74**, 3511–3514 (1995).

Colloidal particles

Origin of anomalous multibody interactions

In a numerical study, Bowen and Sharif¹ purported to capture attractive interactions between like-charged colloidal particles by solving the nonlinear Poisson–Boltzmann equation around two particles in a pore. We are unable to replicate their results, in accord with recent proofs^{2,3} showing that this equation cannot produce attractions between like-charged particles. Our calculations reveal only repulsive interactions between confined particles, indicating that the cause for anomalous colloidal multibody interactions is still unknown.

Colloidal interactions are critical to a variety of processes, including bacterial adhesion⁴, the flow of colloidal dispersions⁵, and the self-assembly of nanoparticles to form optoelectronic devices⁶. Electrostatic interactions between colloidal particles are fairly well described by the Derjaguin–Landau–Verwey–Overbeek (DLVO) theory^{7,8}. However, direct measurement has revealed qualitative departures from this theory in multibody, low-salt situations^{9,10}. Theoretical work^{11–15} has tried to describe the attractions between like-charged colloidal particles.

Like Bowen and Sharif, we solved the nonlinear Poisson–Boltzmann equation in an axisymmetric geometry. We performed finite-element simulations using FIDAP software with successive grid refinement to check the convergence of our results. A fine grid is needed to obtain accurate results,

and integrating the stress tensor over the midplane between particles and over a parallel plane far from the particles allows us to obtain faster convergence than was achieved by Bowen and Sharif, although we checked our results using their technique of integrating over the sphere surface. Our typical grid contains 100,000 nodes, including 220 on the surface of the sphere, twice as many as Bowen and Sharif. We also used regular grid spacings near the important boundaries to minimize numerical noise.

We found that colloidal particles in a pore are repulsive at all distances — indeed, the pore has only a small mitigating effect on the repulsions relative to isolated particles (Fig. 1). Using the same parameters as Bowen and Sharif, the gap between the particle surface and the pore surface is approximately eight Debye lengths, so only a small perturbative effect was expected. (The Debye length, κ^{-1} , is inversely proportional to the square root of the salt concentration, and represents the decay length of screened electrostatic interactions.)

Multibody effects are known to mitigate repulsions between cylinders in planar slits¹¹, but significant effects are not seen until the confining walls are within the range of 1–2 Debye lengths. Increasing the confinement of the pore geometry by reducing the pore wall–particle gap to one Debye length results in significantly smaller repulsive interactions than for isolated spheres (Fig. 1), but there are still no attractions.

Our results are in keeping with proofs^{2,3} that the nonlinear Poisson–Boltzmann equation cannot produce attractive forces for the conditions simulated here and in general for confined or unconfined colloids in a system with mirror symmetry and overall electroneutrality. Together with these proofs, our results indicate that there may be an insidious error in the computations of Bowen and Sharif. In addition, the proofs^{2,3} show that solutions of the mean-field nonlinear Poisson–Boltzmann equation cannot explain electrostatic interactions between like-charged colloidal particles⁹, so new physical theories need to be developed.

Jeffrey J. Gray, Benny Chiang,

Roger T. Bonnecaze

Department of Chemical Engineering,

University of Texas at Austin,

Austin, Texas 78712-1062, USA

1. Bowen, W. R. & Sharif, A. O. *Nature* **393**, 663–665 (1998);

correction *Nature* **402**, 841 (1999).

2. Neu, J. C. *Phys. Rev. Lett.* **82**, 1072–1074 (1999).

3. Sader, J. E. & Chan, D. Y. C. *J. Colloid Interface Sci.* **213**,

268–269 (1999).

4. Razatos, A. P., Ong, Y. L., Sharma, M. M. & Georgiou, G. *Proc.*

Natl Acad. Sci. USA **95**, 11059–11064 (1998).

5. Chen, L.-B., Ackerson, B. J. & Zukoski, C. F. *J. Rheol.* **38**, 193–216 (1994).

6. Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. *Nature* **386**, 143–149 (1997).

7. Derjaguin, B. V. & Landau, L. *Acta Physicochim.* **14**, 633–662 (1941).

8. Verwey, E. J. W. & Overbeek, J. T. G. *Theory of the Stability of Lyophobic Colloids: The Interaction of Sol Particles Having an Electric Double Layer* (Elsevier, New York, 1948).

9. Crocker, J. C. & Grier, D. G. *Phys. Rev. Lett.* **77**, 1897–1900 (1996).

10. Larsen, A. E. & Grier, D. G. *Nature* **385**, 230–233 (1997).

11. Ospeck, M. & Fraden, S. *J. Chem. Phys.* **109**, 9166–9171 (1998).

12. Kwon, G. W., Won, Y. S. & Yoon, B. J. *J. Colloid Interface Sci.*

205, 423–432 (1998).

13. Stevens, M. J., Falk, M. L. & Robbins, M. O. *J. Chem. Phys.* **104**, 5209–5219 (1996).

14. Tehver, R., Ancilotto, F., Toigo, F., Koplik, J. & Banavar, J. R. *Phys. Rev. E* **59**, R1335–R1338 (1999).

15. Goulding, D. & Hansen, J.-P. *Europhys. Lett.* **46**, 407–413 (1999).

corrections

Hiding messages in DNA microdots

C. Taylor Clelland, V. Risco, C. Bancroft

Nature **399**, 533–534 (1999)

In line 16 of the second paragraph, “three-million-fold” should be replaced by “thirty-million-fold”. This correction strengthens our original conclusions. In addition, some of the information in Fig. 1b was either missing or incorrect. The amended figure is shown here.

b

Encryption key

A=CGA	K=AAG	U=CTG	0=ACT
B=CCA	L=TGC	V=CCT	1=ACC
C=GTT	M=TCC	W=CCG	2=TAG
D=TTG	N=TCT	X=CTA	3=GCA
E=GGC	O=GGA	Y=AAA	4=GAG
F=GGT	P=GTG	Z=CTT	5=AGA
G=TTT	Q=AAC	=ATA	6=TTA
H=CGC	R=TCA	=TCG	7=ACA
I=ATG	S=ACG	=GAT	8=AGG
J=AGT	T=TTT	=GCT	9=GCG

Ageing, fitness and neurocognitive function

A. F. Kramer *et al.*

Nature **400**, 418–419 (1999)

The key to Fig. 1b contained two errors: the second set of (dark-blue) bars represents the “incompatible pre-exercise” responses, and the third set of (light-brown) bars represents the “compatible post-exercise responses”.

Dichromatism in macaque monkeys

A. Onishi *et al.*

Nature **402**, 139–140 (1999)

The eighth name in the list of authors, Hidehiko Komatsu, was spelt incorrectly. In the reference list, the first author of ref. 12, Jørgensen, A. L., was also spelt incorrectly.

Parasitoid behaviour and *Bt* plants

T. H. Schuler, R. P. J. Potting, I. Denholm, G. M. Poppy

Nature **400**, 825–826 (1999)

Figure 1b was incorrectly labelled: the bars on both sides should have read “*Bt*-resistant hosts”.

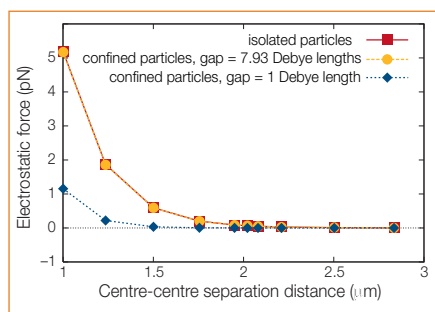


Figure 1 Plot of electrostatic force against distance between particles. Particles are either isolated, confined in a pore of radius $9.12 \kappa^{-1}$, or confined in a pore of radius $2.185 \kappa^{-1}$. For all simulations, the zeta potentials on the pore wall and on the surfaces of the particles are 128.5 and 77.1 mV, respectively, and $\kappa a = 1.185$, where the particle radius $a = 325$ nm. The salt is monovalent, the temperature is 298 K, and the relative permittivity is 78. Parameter values are the same as in ref. 1.

brieF communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com) or on request from nature@nature.com

Probable detection of starlight reflected from the giant planet orbiting τ Boötis

Andrew Collier Cameron*, Keith Horne*, Alan Penny† & David James*

* School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK

† Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK

In the four years following the discovery of a planet orbiting the star 51 Pegasi, about 20 other planets have been detected through their influence on the radial velocities of lines in the stellar spectra. The orbital motion of the planet is detected through the smaller 'reflex motion' of the star, which can be measured using high-precision spectroscopy. This indirect technique cannot investigate the radius or composition of the planet, and can place only a lower limit on its mass. Here we report the probable detection of Doppler-shifted starlight reflected from the planet known to orbit τ Boötis with a period of just a few days. We find that the orbital inclination is about $i = 29^\circ$, from which we infer that the mass is about eight times that of Jupiter. The planet has the size and reflectivity expected for a gas-giant planet.

One of the most surprising features of the planet discovered orbiting 51 Peg was its short orbital period, which indicated that it is very close to its parent star¹—much closer than Jupiter's orbital radius, and far from the region where giant planets are thought to have formed. The observational technique will preferentially find such planets, but their frequency remained a point of controversy until some other planets were also discovered to have surprisingly short periods^{2–4}. The temperatures of planets orbiting so close to their parent stars will obviously be higher than that of Jupiter, raising the question of their sizes (radii).

A planet orbiting a star scatters back into space some of the starlight it receives. To a distant observer, the ratio of the scattered flux f_p from the planet to the direct flux f_* from its star depends on the planet's size and proximity to the star, which determine the amount of light intercepted, and on the scattering properties of the planet's atmosphere. For a planet of radius R_p in an orbit of radius a , the maximum planet-to-star brightness ratio is $\epsilon \equiv f_p/f_* = p(\lambda)(R_p/a)^2$, when the planet is observed from above the sub-stellar point. Here $p(\lambda)$ is the geometric albedo of the atmosphere for light of wavelength λ . The τ Boo planet² has a circular orbit with a radius $a = 0.0462(M_*/1.2M_\odot)^{1/3}$ AU, derived from Kepler's law with the 3.3-day orbital period and estimated stellar mass. Here M_* is the mass of the star, and $M_\odot$ is the mass of the Sun. If this is a giant planet, with a Jupiter-like size and albedo, we expect $\epsilon \sim 10^{-4}$; the scattered starlight will be 10,000 to 20,000 times fainter than the star even when viewed at the most favourable illumination angle $\alpha = 0$.

Although τ Boo and its planet are never separated on the sky by more than 0.003 seconds of arc, this close-in planet orbits its star with a relatively large velocity $V_p = 152 \text{ km s}^{-1}$, based on the star's mass and the planet's orbital period². The star and planet may therefore be easier to separate by using the orbital Doppler shift than by direct imaging. The pattern of photospheric absorption lines in the starlight is preserved when the starlight is reflected from the planet, apart from the Doppler shift due to the planet's orbital velocity, and a multiplicative scaling by the wavelength-dependent albedo of its atmosphere. If the planet's orbit is inclined by an angle i relative to our line of sight, its Doppler shift varies between $-K_p$ and $+K_p$ as it orbits around the star, where $K_p = V_p \sin i$. For comparison, the star's absorption lines span $\pm 28 \text{ km s}^{-1}$, largely due to the star's rotation. Except in the unlikely case of a nearly face-on orbit, $i < 10^\circ$, the orbital Doppler shift should cleanly separate the planet's spectral lines from those of the star.

Observations of τ Boo

We observed τ Boo for a total of 48 hours on four nights in April

1998 and five further nights in April, May and June 1999 using the Utrecht Echelle Spectrograph on the 4.2-m William Herschel Telescope (WHT) at the Roque de los Muchachos Observatory on La Palma. The ranges in orbital phase covered by each night's observations are listed in Table 1.

Separating direct and reflected spectral lines

To isolate the feeble scattered-light spectrum of τ Boo's planet, we begin by constructing a high-fidelity spectral model to accurately subtract the direct starlight from each of the 580 spectra. The initial stellar spectrum subtraction is a delicate operation. Small shifts of the spectrum on the detector and changes in the telescope focus and atmospheric seeing distort the spectral-line profiles from each exposure to the next. Even small distortions of the strong stellar lines can produce changes larger than the faint signature of the planet. We therefore calibrate these instrumental distortions by comparing each spectrum with a 'template' spectrum $T(\lambda)$, which we construct by adding together all other spectra secured on the same night. Making a new template for each night, rather than a single template for the whole observing run, helps to remove small changes in the detector response pattern from one night to the next. The stability of the spectrograph is such that all spectra secured on a given night are mutually aligned to within 0.2 pixel (0.6 km s^{-1}), so that the template spectra are not appreciably broadened. Our template spectra have a signal-to-noise ratio exceeding 10^4 per spectral pixel in the best-exposed parts of the image.

We model each of the individually distorted stellar spectra as a linear combination of the template and its first and second derivatives with respect to wavelength:

$$S(\lambda, \phi) = a_0 T(\lambda) + a_1 \frac{\partial T(\lambda)}{\partial \lambda} + a_2 \frac{\partial^2 T(\lambda)}{\partial \lambda^2}$$

The line spread function (LSF) in our spectra is only 2.1 pixels wide (6.4 km s^{-1}), but the star's rotation spreads the stellar line profiles across 18 pixels (55 km s^{-1}). The above approximation is therefore adequate (in terms of stability against 'ringing' and other related sampling effects) to correct the 0.2-pixel shifts and ~ 0.5 -pixel seeing-induced changes we see in the width of the LSF. The coefficient a_0 serves to scale the template to match the observed spectrum, a_1 models the pixel shift between the spectrum and the template, and a_2 accounts for small changes in the width of the LSF caused primarily by changes in atmospheric seeing. We at first included higher derivatives to model changes in the skew and kurtosis of the LSF, but concluded that these effects were not significant. To account for small changes in the LSF of the spectro-

graph across the focal plane, we represent the three coefficients as spline functions that vary smoothly across the frame. We optimize the spline functions to fit each spectrum, using an iterative least-squares fit with outlier rejection, and then subtract the optimized model from each observed spectrum in turn.

This procedure very effectively removes the direct starlight signal, leaving the planet signal deeply buried in noise. The procedure of constructing a template spectrum to model the direct starlight spectrum inevitably incorporates into the template some part of the planet signal. This is particularly so near the quadrature phases 0.25 and 0.75, when the planet's velocity is approximately constant. Subtracting the template thus modifies and eliminates some of the planet signal. We account for this partial suppression later in the signal extraction process.

After subtracting the template to remove the stellar spectrum, we add together groups of four contiguous observations, primarily to save computer time in the subsequent analysis. This reduces the number of spectra from 580 to 145, and boosts the signal-to-noise ratio from ~ 600 to $\sim 1,200$ per spectral pixel. The planet signal, expected to be at least 10^4 times fainter than the starlight, remains deeply buried in noise at this stage of the analysis.

The planetary scattered-light signature is present in each spectrum as a faint Doppler-shifted copy of each of the star's spectral lines. To build up this signal, we combine the velocity profiles of thousands of spectral lines using a least-squares deconvolution (LSD) technique that has become a standard tool for detecting and mapping stellar surface features⁵.

Our LSD procedure uses a list of the known wavelengths and strengths of $\sim 2,300$ spectral lines that are present over the observed wavelength domain in a star of τ Boo's temperature and composition. We determine, using χ^2 minimization, a 'mean' line profile which, when convolved with the known line pattern, yields an optimally weighted match to the strengths and shapes of the lines in the observed spectrum. This 'deconvolved velocity profile' is thus an average that is representative of all the lines recorded in the spectrum, but with a vastly improved signal-to-noise ratio. Blends are automatically compensated for, so that LSD gives a flat continuum outside the stellar profile, free of the sidelobes that are produced by simpler shift-and-add or cross-correlation procedures. We first apply LSD to the spectra of τ Boo, to establish the mean strength of its stellar absorption lines. A second application of LSD, this time to the template-subtracted spectra, extracts the mean velocity profile of the planetary signal averaged over all of those lines. The resulting template-subtracted velocity profiles, one for each spectrum, are displayed as a velocity-phase map in Fig. 1.

The most prominent pattern visible in Fig. 1 is a 'barber's pole' pattern at low velocities inside $\pm 28 \text{ km s}^{-1}$. This pattern is the same when we split the data into several wavelength regions, suggesting a stellar origin rather than a problem in the template registration and

subtraction. The ripples drift from the blue to the red wing of the profile in roughly half the orbital period of the planet. One explanation is the possible presence of solar-sized starspot groups crossing the face of the star as it rotates synchronously with the orbit. The presence of large-scale inhomogeneities in the photospheric velocity field⁶ could also give low-amplitude, rotationally-modulated distortion of the line profiles without producing significant optical variability. A rotational modulation of the star's chromospheric Ca II H and K emission has been detected with a period of 3.3 days (ref. 7), essentially the same as the 3.312567-day orbital period of the planet. The barber's-pole pattern thus provides useful independent confirmation that the star does indeed rotate synchronously with the planet's orbit.

Outside the range $\pm 28 \text{ km s}^{-1}$, the residuals have the general appearance of random noise, but with some correlation in time due to nights with residual flat-fielding errors. The root-mean-square scatter indicates a signal-to-noise ratio of $\sim 3.5 \times 10^4$ per spectral pixel relative to the original stellar continuum level. This noise level is 10% to 15% greater than photon noise estimates propagated through the data extraction and deconvolution. LSD has therefore realized a 30-fold increase in signal-to-noise ratio by combining the velocity profiles of the $\sim 2,300$ spectral lines. This should be just sufficient to detect the brightest expected planet signature in those of the 145 LSD velocity profiles that are taken at gibbous phases when the planet is Doppler-shifted well clear of the stellar profile.

To search for candidate planet signals, measure their strengths and assess their significance, and to quantify upper limits on planets at other inclinations, we must extract information from the LSD velocity profiles at all 145 orbital phases. We do this by using a matched filter to account for the expected changes in the planet's

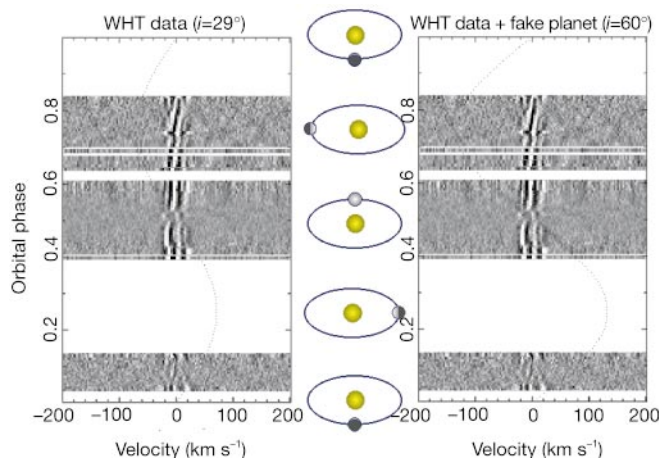


Figure 1 Greyscale plots showing 145 residual velocity profiles. The profiles arise from subtracting the stellar template model and averaging the profiles of $\sim 2,300$ spectral lines in each of the 145 spectra. The velocity scale is in the reference frame of the star, and time increases upward. The orientation of the system at each orbital phase is sketched at centre. The greyscale runs from black at -10^{-4} to white at $+10^{-4}$ times the mean stellar continuum flux. The dotted paths indicate the velocity curve of a planet orbiting with an orbital inclination of 29° (left-hand panel) and 60° (right-hand panel). The right-hand panel shows the effect of adding the simulated spectrum of a planet with 1.4 Jupiter radii and a geometric albedo of 0.55 to the original spectra. The signature of this simulated planet appears as a dark linear feature crossing from $+80 \text{ km s}^{-1}$ at phase 0.4 to -80 km s^{-1} at phase 0.6. The planetary signature detected in the data is much weaker, because of the low inclination, but would follow the velocity curve shown in the left-hand panel. The 'barber's pole' pattern of travelling ripples lies wholly within the residual stellar profile and has a characteristic amplitude of $(2-4) \times 10^{-4}$ of the mean stellar continuum level. Its form remains invariant in profiles deconvolved from independent subsets of the data at different wavelengths. All data between phases 0.64 and 0.83 were obtained in a five-night interval during the 1998 season. The ripple pattern appears both stronger and more coherent here than in the 1999 dataset, with five nights spread over two months selecting orbital phases near 0.5.

Table 1 Journal of observations

UTC start	Phase	UTC end	Phase	Number of groups
1998 Apr 09 22:09:43	0.4254	1998 Apr 10 05:37:05	0.5192	25
1998 Apr 10 22:04:40	0.7262	1998 Apr 11 06:20:54	0.8303	29
1998 Apr 11 22:09:28	0.0291	1998 Apr 12 05:58:44	0.1275	21
1998 Apr 13 23:02:54	0.6441	1998 Apr 14 06:08:23	0.7332	9
1999 Apr 02 22:06:45	0.4931	1999 Apr 03 06:08:47	0.5941	15
1999 Apr 25 21:43:17	0.4411	1999 Apr 26 05:31:59	0.5393	10
1999 May 05 21:56:59	0.4628	1999 May 06 04:51:57	0.5497	15
1999 May 25 20:59:45	0.4884	1999 May 26 03:56:18	0.5757	11
1999 May 28 21:06:19	0.3954	1999 May 29 02:40:44	0.4655	10

The UTC mid-times and orbital phases are shown for the first and last groups of four spectra secured on each night of observation. The spectra spanned the optical wavelength range from 385 nm to 611 nm at a resolving power $\lambda/\Delta\lambda = 49,000$. The signal-to-noise ratio, typically ~ 600 in each spectral pixel of width 3 km s^{-1} , was limited by photon statistics. The number of groups of four such consecutive spectra is given in the final column. The orbital phase ϕ increases from 0 to 1 around the orbit, and is defined so that the planet is closest to the observer at phase 0. In the 1999 season we concentrated on the phase range between 0.4 and 0.6, where the planet is expected to be brightest. The orbital phases are computed from the epoch of inferior conjunction T_0 at HJD 2451269.756 and orbital period $P = 3.312567$ days (G. Marcy, personal communication).

Doppler shift and brightness with orbital phase.

We model the velocity profile of a planet signal as a moving gaussian:

$$G(v, \phi, K_p) = \frac{Wg(\phi, i)}{\sqrt{\pi}\Delta} \exp\left\{-\left(\frac{v - K_p \sin \phi}{\Delta}\right)^2\right\}$$

The velocity width parameter Δ is chosen to match the expected averaged width of the stellar lines reflected from the planet. Because the star rotates synchronously with the planet's orbit, the planet sees the stellar spectrum without the rotational broadening that we see in the direct starlight⁸. We estimate the width parameters of the direct starlight and the reflected light to be 13.2 and 6.4 km s⁻¹ respectively, from gaussian fits to the deconvolved profiles of τ Boo and the slowly rotating giant star HR 5694. The latter has an F7 III spectral type (temperature) and elemental abundances similar to τ Boo (ref. 7). The larger width and shallower depth of τ Boo's lines, due largely to the star's axial rotation, yields⁹⁻¹¹ a projected equatorial rotation speed $v \sin i = 14.8 \pm 0.3$ km s⁻¹, where i is the inclination of the spin axis to the line of sight.

The line strength parameter W is set to match that of τ Boo. For $g(\phi, i)$, which modulates the planet's brightness with orbital phase, we adopt an empirically determined polynomial approximation to the phase function of Venus¹². The orbital velocity amplitude is $K_p = V_p \sin i$, where $V_p = 152$ km s⁻¹. Note that K_p and i are not independent parameters; we compute i for each value of K_p .

To compensate for the attenuating effect of a blurred planet signal being present in the template, we mimic the effect in constructing the matched filter $M(v, \phi, K_p)$. We do this by subtracting the flux-weighted average of the gaussian planet signals on each night, emulating precisely the attenuation arising from the template construction and subtraction.

To allow for the possibility that the planet's albedo is wavelength-dependent, we subdivide our data into six independent wavelength ranges and measure the planet signal strength $\epsilon(\lambda)$ in each data subset. Thus we construct template-subtracted LSD velocity profiles $f(v, \phi, \lambda)$, and their corresponding noise variances $\sigma^2(v, \phi, \lambda)$, for 6

independent subsets of the echelle orders at wavelengths λ . For any set of trial values of the parameters K_p and $\epsilon(\lambda)$, we quantify the 'badness of fit' to the data by means of the standard χ^2 statistic:

$$\chi^2 \equiv \sum_v \sum_\phi \sum_\lambda \left(\frac{f(v, \phi, \lambda) - \epsilon(\lambda)M(v, \phi, K_p)}{\sigma(v, \phi, \lambda)} \right)^2$$

Thus we scale the matched filter $M(v, \phi, K_p)$ by factors $\epsilon(\lambda)$ to fit LSD velocity profiles $f(v, \phi, \lambda) \pm \sigma(v, \phi, \lambda)$, which are like those shown in Fig. 1 but constructed from wavelength-restricted subsets of the data. The best fit minimizes χ^2 , yielding the optimal estimates of K_p and $\epsilon(\lambda)$. The increase in χ^2 for nearby parameter values, $\Delta\chi^2$, is used to judge different models, their relative probabilities being proportional to $\exp(-\Delta\chi^2/2)$. To prevent spurious planet signatures arising from the barber's-pole pattern in the residual stellar profile, we exclude from the fitting procedure all pixels within 31.5 km s⁻¹ of the centre of the stellar profile.

Testing the detection procedure

We verified that our procedure is capable of revealing faint planetary signals in the presence of realistic noise levels and patterns by adding a simulated planetary signal to the observed spectra, then repeating the template construction, subtraction and LSD analysis. The planet was simulated by the spectrum of HR 5694, Doppler-shifted to the appropriate orbital velocities and scaled according to the phase function expected for a planet with a radius 1.4 times that of Jupiter and a grey albedo $p = 0.55$ (Fig. 1).

In searching for candidate planetary signals in the WHT data, we begin by testing the null hypothesis that no planet is present, in comparison with the alternative hypothesis that a 'grey' planet with wavelength-independent albedo is present at some value of K_p . We probe for evidence of a planet over the range $40 \text{ km s}^{-1} < K_p < 152 \text{ km s}^{-1}$, corresponding to inclinations $15^\circ < i < 90^\circ$. At lower velocities and inclinations the planet's orbital velocity never emerges from the low-velocity region affected by the barber's-pole pattern. An inclination less than $i = 15^\circ$ would also require an implausibly high stellar equatorial rotation speed¹³.

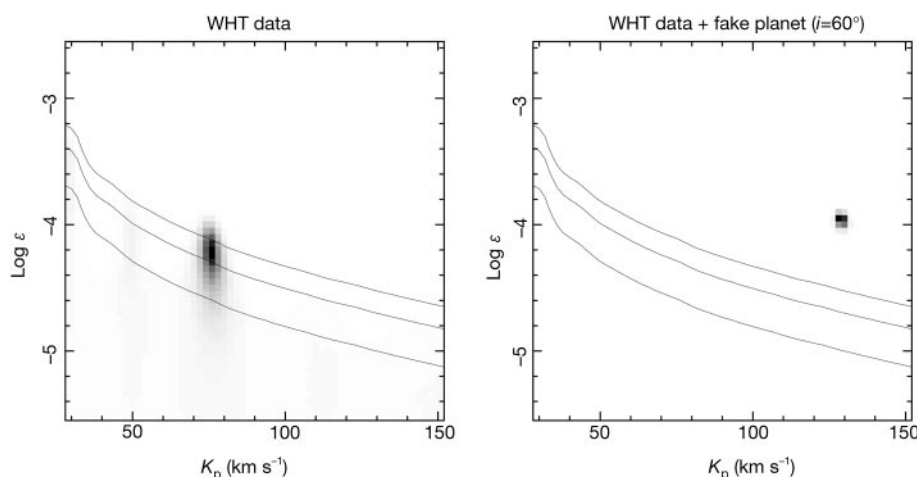


Figure 2 Improvement in the fit to the data over the no-planet hypothesis, shown for the simulation of a planet with a wavelength-independent (grey) albedo. The two panels present identical analyses of the residual velocity profiles shown in the corresponding panels of Fig. 1, representing the William Herschel Telescope (WHT) data alone (left) and the WHT data after adding a simulated planet signal for $i = 60^\circ$ (right). For each possible value of the unknown planet-to-star brightness ratio ϵ , and of the planet's projected orbit velocity K_p , darker shades denote progressively better fits to the data. The probability relative to the best-fit model increases from 0 for white to 1 for black. In the right panel, the fake planet's parameters $K_p = 132$ km s⁻¹ and $\epsilon = 1.07 \times 10^{-4}$ are correctly recovered. In the left panel, the best fit to the WHT data occurs for $K_p = 74$ km s⁻¹ and $\epsilon = 7.5 \times 10^{-5}$. The 'grey planet' model significantly improves the fit (darker shading)

relative to the 'no planet' model ($\epsilon = 0$); the improvement in χ^2 after optimizing ϵ and K_p is $\Delta\chi^2 = 9.74$. If no planet were present, the probability of such an improvement occurring by chance should have a χ^2 distribution with two degrees of freedom. The actual distribution, derived for the noise and systematic error in our data via a bootstrap error analysis, has a slightly longer tail, so that $\Delta\chi^2 = 9.74$ rejects the 'no planet' hypothesis with 97.8% confidence. If no planet is present, the best-fit model should lie below the 68.3%, 95.4%, and 99.7% bootstrap confidence limits shown as solid curves. These limits decrease with K_p because for larger K_p the planet lines move more rapidly across the stellar profile and into the region covered by the data between phases 0.4 and 0.6. The light grey shading below the confidence limits is consistent with noise, and planets are strongly excluded in the white zones above the highest confidence limit.

Figure 2 presents the relative-probability map as a function of K_p and ϵ , showing significant evidence for a planet at $K_p = 74 \pm 3 \text{ km s}^{-1}$ and $\epsilon = 7.5 \pm 3 \times 10^{-5}$. The ‘grey planet’ hypothesis introduces two parameters: ϵ and K_p . The improvement in the fit, reducing χ^2 by $\Delta\chi^2 = 9.74$ with only two degrees of freedom, rejects the ‘no planet’ hypothesis with 97.8% confidence, based on a bootstrap analysis of the error distribution. Thus, if no planet is present, the probability is about 2% that the noise in our data would produce a spurious signal this strong. When the synthetic planet signature at $i = 60^\circ$ is injected into the data, its velocity amplitude and strength are recovered correctly (Fig. 2).

We examine next the hypothesis that the albedo depends on wavelength, by testing the ‘no planet’ hypothesis against a model with seven additional parameters: K_p , and $\epsilon(\lambda)$ for six wavelengths λ . Thus we split the data into six independent wavelength bands chosen so that the recorded flux is divided roughly equally among the six bands. By fitting independent values of ϵ to different wavelength ranges we allow for the possibility that the planet’s albedo spectrum could contain broad absorption bands due to molecular or atomic species such as TiO or Na I overlying the main scattering layers. The greatest improvement in χ^2 occurs again at $K_p = 74 \text{ km s}^{-1}$ (Fig. 3). With $\Delta\chi^2 = 37.96$, the ‘no planet’ hypothesis is again rejected, this time with 99.2% confidence for seven degrees of freedom, a 0.8% probability of occurring by chance.

This result implies a significant departure from a wavelength-independent albedo spectrum. The signal is concentrated in the three wavelength bands from 456 nm to 524 nm, with the central 479–509-nm band contributing half of the signal. Averaging over these three bands alone, we obtain $\epsilon = (1.9 \pm 0.4) \times 10^{-4}$. No significant signal is present in the other bands redward and blueward of these three.

We conclude that there is strong evidence in our data for a planetary signal at $K_p = 74 \text{ km s}^{-1}$, which cannot easily be explained as a spurious detection caused by the photon statistics and locally correlated errors in our data.

The results obtained so far have relied on calculations of the planet’s orbital phase based on an orbit period $P = 3.312567$ days and conjunction epoch T_0 (G. Marcy, personal communication). These parameters were derived from high-precision studies of the star’s velocity, and we therefore hold P and T_0 fixed at their known values while searching for reflected-light signals over the feasible range of the unknown orbit velocities K_p . If the planet signal is strong enough, however, we should be able to measure P and T_0 independently from our data, and those values should be consistent with the known values. Considering a small range around Marcy’s values, our data yield $P = 3.3128 \pm 0.0002$ days (1 σ errors), consistent with Marcy’s more accurate period, and a conjunction phase $\phi_0 = 0.0007 \pm 0.003$, consistent with Marcy’s 0.000 ± 0.002 . Our measurement of $K_p = 74 \pm 2 \text{ km s}^{-1}$ is also insensitive to small changes in the period and epoch of the planet’s orbit.

We also considered different values for the velocity width Δ of the absorption lines in the light scattered from the planet. The best fit was found for $\Delta = 6 \pm 2 \text{ km s}^{-1}$. This is in good agreement with the 6.4 km s^{-1} width of the lines in the spectrum of HR 5694, confirming that the planet sees the starlight without rotational broadening, and hence that the star rotates synchronously with the planet.

As an additional test, we use a periodogram analysis to estimate the probability that our planet signal could be an artefact of noise. For this test, we relax our knowledge of the planet’s orbital period P , and re-analyse the data for many different trial periods in the range $3.25 < P < 3.35$ days. We estimate the ‘false alarm’ probability as the fraction of periods that yield a planet signal stronger than our candidate detection for some value of K_p in the feasible range. In order to keep the orbital phases close to 0.5 in the 1999 data for all periods considered, we set the time of zero phase to an epoch of

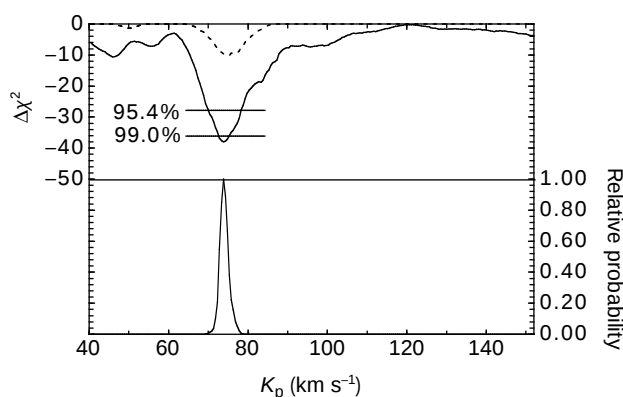


Figure 3 Evidence for a planet in the WHT data, assessed as a function of the planet’s orbit velocity K_p . The top panel gives $\Delta\chi^2$, the reduction in the χ^2 of the fit relative to the assumption that no planet is present, for the two-parameter model in which the planet is assumed to have constant (grey) albedo (dashed curve), and for the seven-parameter model with the planet’s albedo varying with wavelength (solid curve). The horizontal lines give, for two values of $\Delta\chi^2$, the corresponding probabilities that the improvement in χ^2 is too large to be attributed to noise. The two-parameter constant-albedo model gives a significant (97.8%) improvement at $K_p = 74 \text{ km s}^{-1}$, and the wavelength-varying albedo model with seven fitted parameters raises this probability to 99.2%. The lower panel shows the probability of the seven-parameter model as a function of K_p , scaled to 1 for the best-fit model $K_p = 74 \pm 2 \text{ km s}^{-1}$.

conjunction just before 5 May 1999. The pattern of spurious peaks as a function of both K_p and P is then as shown in Fig. 4. The lower panel of this figure shows the minimum value of χ^2 , optimized over the range $40 < K_p < 152 \text{ km s}^{-1}$, at each period. Of all the periods sampled, a small fraction yield spurious detections stronger than our planet signal, and close enough to their local maxima to pass the same local tests as our candidate signal.

We conclude from the periodogram test that the probability of encountering a spurious peak with $K_p > 74 \text{ km s}^{-1}$ is between 3% and 5%. This is our most pessimistic assessment of the probability that our detection is spurious. If other prior knowledge is taken into account—in particular, the restrictions imposed on the orbital inclination by the star’s synchronous rotation—the false-alarm probability is reduced further still.

Deriving the planet’s properties

The best-fitting orbital velocity amplitude $K_p = 74 \text{ km s}^{-1}$ yields an orbital inclination $i = 29^\circ$ and implies a planet mass eight times that of Jupiter, twice the minimum value for an edge-on orbit. The low inclination is consistent with arguments for synchronous rotation, based on the star’s equatorial rotation speed, rotation period, angular diameter and distance^{7,10}, plus the requirement that the star’s rotational synchronisation timescale ($1.2 \times 10^{11} \sin^8 i$ years) must then be shorter than its present age (less than ~ 3 Gyr) (refs 14, 15).

At this low inclination, the fact that we are able to detect the planet at all suggests a large radius and/or reflectivity. Indeed, our planet-to-star flux ratio $\epsilon = (1.9 \pm 0.4) \times 10^{-4}$, averaged between 456 and 524 nm, appears to conflict with recent work¹³ that establishes an upper limit on the brightness of the τ Boo planet $\epsilon < 10^{-4}$ with 99.9% confidence for $i \sim 30^\circ$ at similar wavelengths. We caution, however, that the value of ϵ produced by our fitting method is sensitive to the Venus-like form we adopt for the phase function $g(\phi, i)$. A lower value of ϵ and hence a smaller radius results if the true phase function is less strongly peaked at the zero phase angle than the one used here. We also note that, while Charbonneau *et al.*¹³ estimate a $\sim 20\%$ attenuation of planet signals due to their template-subtraction process, their tests recovering simulated

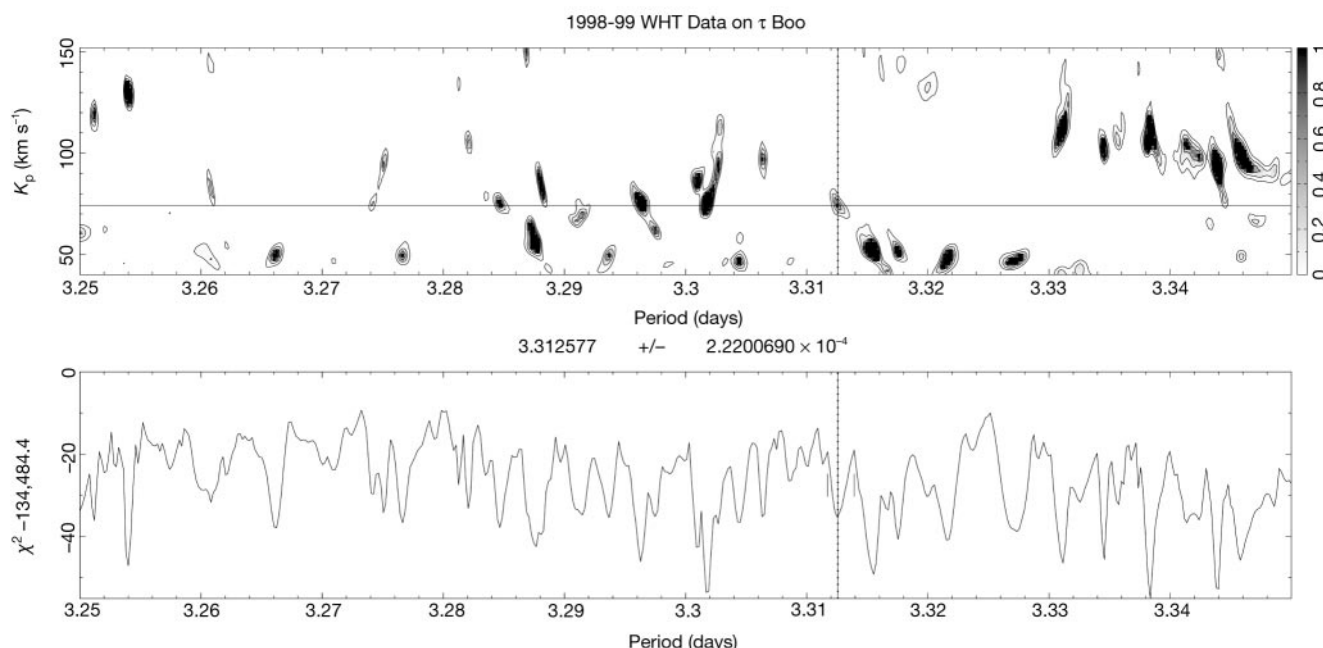


Figure 4 This periodogram analysis searches for evidence of reflected light from the τ Boo planet over a range of trail values of the orbit velocity K_p and orbit period P . For each value of K_p and P , the six values $\epsilon(\lambda)$ are optimized to fit six independent wavelength subsets of the echelle data. The lower panel shows the best χ^2 found for K_p between 40 and 152 km s⁻¹ at each trial period P . In the upper panel, the greyscale indicates for each value of K_p and P the relative probability of the best-fit model, with white to black representing increasing relative probability. The vertical line marks the accurate value of P , from analysis of the star's Doppler signal (G. Marcy, personal communication). Our detection is the probability peak occurring at the correct period near $K_p = 74$ km s⁻¹, as

marked by the intersection between the vertical and horizontal lines. The contours give 1σ , 2σ , and 3σ confidence regions for this peak, based on $\Delta\chi^2$ for two degrees of freedom. A number of other peaks, some giving even better fits to the data than our detection, are seen at different values of P and K_p . These peaks arise mainly from noise in the data. We infer from the number of such peaks that there is a 3% to 5% probability that our detection is a spurious noise peak that happens by chance to coincide with the known period and phase. Note that some periodogram features appear twice, shifted horizontally by 0.03 days. These correspond to pairs of periods that differ by one orbital cycle per year, due to the one-year gap that separates our 1998 and 1999 observations.

planet signatures were incomplete because the template spectrum was fitted before rather than after adding the planet. We designed our matched-filter fitting function to mimic closely the effects of template attenuation. Our tests with recovery of synthetic planet signatures indicate that this was adequate to compensate for the effect. We conclude that differences in the treatment of the phase function and template attenuation effects are sufficient to reconcile the apparent conflict between our detection and the upper limit published by Charbonneau *et al.*¹³

The radius we infer for τ Boo's planet is 1.8 times that of Jupiter, assuming a Jupiter-like albedo $p = 0.55$, and using the Venus-like phase function. This is somewhat larger than the radii predicted by recent structural and evolutionary models¹⁶, which range from $\sim 1.4R_J$ at age 2 Gyr to $\sim 1.1R_J$ at age 3 Gyr, where R_J is Jupiter's radius. If we use a Lambert-sphere model¹³, our inferred radius decreases by 12% to $1.6R_J$.

Our candidate detection of starlight scattered from the atmosphere of an extrasolar planet strengthens the case for the existence of the giant, close-orbiting planets whose presence has so far been inferred only indirectly from the reflex motions of their parent stars. The strength of the detection indicates that the τ Boo planet is a gas giant, and suggests that its optical reflectivity may be appreciable only over a narrow range of blue-green wavelengths. Further observations will be needed to determine the form of the phase function, and so improve estimates of the radius and albedo. We envisage that it will be possible to infer the presence or absence of molecular species such as TiO, methane and water in the vapour phase, by carrying out more sensitive studies over suitably defined wavelength ranges. Finally, we note that the close-orbiting planets of other systems, including both 51 Pegasi and the ν Andromedae triple-planet system, should be amenable to similar studies in the near future. □

Received 6 July; accepted 16 November 1999.

- Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359 (1995).
- Butler, R. P., Marcy, G. W., Williams, E., Hauser, H. & Shirts, P. Three new 51 Pegasi-type planets. *Astrophys. J.* **474**, L115–L118 (1997).
- Butler, R. P., Marcy, G. W., Vogt, S. S. & Aps, K. A planet with a 3.1 day period around a solar twin. *Publ. Astron. Soc. Pacif.* **110**, 1389–1393 (1998).
- Fischer, D. A., Marcy, G. W., Butler, R. P., Vogt, S. S. & Apos, K. Planetary companions around two solar-type stars: HD 195019 and HD 217107. *Publ. Astron. Soc. Pacif.* **111**, 50–56 (1999).
- Donati, J.-F., Semel, M., Carter, B., Rees, D. E. & Collier Cameron, A. Spectropolarimetric observations of active stars. *Mon. Not. R. Astron. Soc.* **291**, 658–682 (1997).
- Toner, C. G. & Gray, D. F. The starpatch on the G8 dwarf ξ Bootis A. *Astrophys. J.* **334**, 1008–1020 (1998).
- Baliunas, S. L., Henry, G. W., Donahue, R. A., Fekel, F. C. & Soon, W. H. Properties of sun-like stars with planets: ρ^1 Cancri, τ Bootis, and ν Andromedae. *Astrophys. J.* **474**, L119–L122 (1997).
- Charbonneau, D., Jha, S. & Noyes, R. W. Spectral line distortions in the presence of a close-in planet. *Astrophys. J.* **507**, L153–L156 (1998).
- Gray, D. F. Observational evidence against differential rotation in F stars. *Astrophys. J.* **258**, 201–208 (1982).
- Fuhrmann, K., Pfeiffer, M. J. & Bernkopf, J. F- and G-type stars with planetary companions: ν Andromedae, ρ^1 Cancri, τ Bootis, 16 Cygni and ρ Coronae Borealis. *Astron. Astrophys.* **336**, 942–952 (1998).
- Gonzalez, G. Spectroscopic analyses of the parent stars of extrasolar planetary system candidates. *Astron. Astrophys.* **334**, 221–238 (1998).
- Hilton, J. L. in *Explanatory Supplement to the Astronomical Almanac* (ed. Seidelmann, P. K.) 383–419 (University Science Books, Mill Valley, California, 1992).
- Charbonneau, D. *et al.* An upper limit on the reflected light from the planet orbiting the star τ Bootis. *Astrophys. J.* **522**, L145–L149 (1999).
- Marcy, G. W. *et al.* The planet around 51 Pegasi. *Astrophys. J.* **481**, 926–935 (1997).
- Drake, S. A., Pravdo, S. H., Angelini, L. & Stern, R. A. Synchronization timescales for three solar-type stars that have Jupiter-mass companions in short-period orbits. *Astron. J.* **115**, 2122–2124 (1998).
- Guillot, T. *et al.* in *IAU Colloquium 161: Astronomical and Biochemical Origins and the Search for Life in the Universe* (eds Cosmovici, C. V., Bowyer, S. & Wertheimer, D.) 343–350 (Editrice Compositori, Bologna, 1997).

Acknowledgements

This work is based on observations with the 4.2-m William Herschel Telescope at the Observatori del Roque de los Muchachos on La Palma, and used Starlink-supported hardware and software.

Correspondence and requests for materials should be addressed to A.C.C. (e-mail: andrew.cameron@st-and.ac.uk).

Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle

Anne Davaille

Laboratoire de Dynamique des Systèmes Géologiques, IPG, 4 Place Jussieu, 75252 Paris cedex 05, France

Mounting evidence indicates that the Earth's mantle is chemically heterogeneous. To understand the forms that convection might take in such a mantle, I have conducted laboratory experiments on thermochemical convection in a fluid with stratified density and viscosity. For intermediate density contrasts, a 'doming' regime of convection is observed, in which hot domes oscillate vertically through the whole layer while thin tubular plumes rise from their upper surfaces. These plumes could be responsible for the 'hot spots' and the domes themselves for the 'superswells' observed at the Earth's surface. In the Earth's mantle, the doming regime should occur for density contrasts less than about 1%. Moreover, quantitative scaling laws derived from the experiments show that the mantle might have evolved from strictly stratified convection 4 Gyr ago to doming today. Thermochemical convection can thus reconcile the survival of geochemically distinct reservoirs with the small amplitude of present-day density heterogeneities inferred from seismology and mineral physics.

The pattern of convection beneath the Earth's surface plates remains one of the main unsolved problems in geophysics¹. Because both seismic Earth models² and mineral physics studies³ suggest that any present-day chemical density heterogeneities in the mantle cannot exceed 1–2% (ref. 4), the mantle is often assumed to be homogeneous for purposes of dynamical modelling. But thermal convection in a homogeneous mantle fails to explain a number of observations: isotopically distinct reservoirs have persisted in the mantle for billions of years (ref. 5); at the base of the mantle, large seismic velocity anomalies indicate a chemical and/or partial melt component⁶; and although some subducted plates have been imaged down to the core–mantle boundary⁷, changes in the tomographic image at mid-mantle to lower-mantle depths suggest a reduced mass transfer there⁸. Furthermore, the origins of two of the principal manifestations of terrestrial intraplate volcanism, 'hot-spots' and 'superswells' remain enigmatic.

Most 'hotspots'⁹ are believed to occur where a hot and narrow (<100 km) plume rising from a deep mantle boundary layer impinges on the lithosphere^{10–12}. Hotspots can remain active for more than 120 Myr, and because they move laterally much more slowly than the lithospheric plates¹², they produce volcanic tracks like the Emperor–Hawaiian chain. Hotspots sample mantle geochemical reservoirs distinct from those sampled by mid-ocean ridges⁵, and their global distribution may influence core dynamics^{13,14}. 'Superswells'¹⁵ are regions of anomalously shallow sea-floor, or high topography, several thousand kilometres in extent with unusually dense concentrations of hotspots. At present, two have been identified: one under Africa¹⁶ and one beneath French Polynesia in the South Pacific¹⁷. Seismic tomography suggests that the mantle beneath these superswells is anomalously hot, and also possibly chemically distinct close to the core–mantle boundary^{15,18}. Furthermore, the lavas produced are radiogenically enriched¹⁹. Modelling based on correlation of seismic tomography, geoid and topography concludes that the Polynesian superswell is caused by dynamic upwelling of hot material in the upper mantle^{15,20} or in the transition zone at around 670 km depth²¹. It has been suggested that such 'superplumes' are oscillatory phenomena^{15,22}, related to surges in ocean-crust production and periods of magnetic quiescence²². If true, this indicates strong coupling of the core and the mantle, and

suggests that superplumes originate at the core–mantle boundary¹⁴.

Phenomena like those just described are difficult to explain in terms of convection in homogeneous viscous fluids. Thermal convection in such fluids is controlled mainly by the Rayleigh number Ra , the ratio of the driving buoyancy forces to the resisting effects of thermal diffusion and viscous dissipation:

$$Ra = \frac{\alpha g \Delta T d^3}{\kappa \nu} \quad (1)$$

where d is the layer depth, ΔT is the temperature drop across it, g is the gravitational acceleration, α is the thermal expansivity, κ is the thermal diffusivity and ν is the kinematic viscosity. For $Ra > 10^6$, as in the mantle, oscillatory superplumes are never observed. But upwellings do have the form of cylindrical plumes²³, which seems at first sight to explain hotspots. However, many mantle plumes are too cold to have arisen from a purely thermal core–mantle boundary layer having the expected temperature jump ($\sim 1,000^\circ\text{C}$)^{24,25}, and others are too weak to have traversed the entire depth of the mantle²⁴. Moreover, mixing of passive heterogeneities in homogeneous convection is too rapid to preserve 'primitive' mantle as required by geochemical data^{26–28}. To slow mixing, viscosity heterogeneities^{29–31} or chemical stratification³² are required, and most probably a combination of both in the case of the mantle³³.

Here I present results from laboratory experiments on thermal convection in a fluid with stratified density and viscosity³⁴. I shall show that even small chemical heterogeneities can completely change the style of convection when the material is heterogeneous in viscosity as well as in density. Weak density stratification generates hot domes which oscillate vertically through the whole layer while thin tubular plumes rise from their upper surfaces. Strong stratification produces two-layer convection with hot thin tubular plumes issuing from cusps on the interface. I suggest that the plumes could be responsible for hotspots, and the domes for superswells. In the mantle, the doming regime will occur for density contrasts less than about 1%. Moreover, quantitative scaling laws derived from the experiments show that the mantle is likely to have evolved from strictly stratified convection 4 Gyr ago to doming today.

Thermal convection in a stratified fluid

I performed laboratory experiments in which two layers (one above the other) of viscous fluids, initially isothermal, were suddenly cooled from above and heated from below. The fluids were mixtures of water, salt and cellulose whose concentrations were independently varied to obtain initial density differences between 0.10% and 7% and viscosity ratios $\gamma (= \nu_2/\nu_1)$ between 10^{-3} and 6×10^4 . Layer '2' is defined as the layer with the higher Rayleigh number. The liquids are miscible in all proportions, and the temperature dependence of the viscosity is negligible (it only decreases by a factor of 4 from 5 to 50 °C) compared to its composition dependence³⁴. The initial density stratification was either sharp or gradual; layer depth ratios ranged between 1 and 30, and Rayleigh numbers (based on the thickness and viscosity of each layer) ranged between 4 and

2×10^8 , with at least one layer having a Rayleigh number greater than 10^6 . The most viscous layer was always the lower. The high viscosities render diffusion of salt extremely slow, so that entrainment of material across the interface is purely mechanical³⁴. Heat and mass transfer were monitored through time by measuring temperature profiles and the densities of both layers.

In all the experiments, convection involved both a small-scale thermal mode (small and short-lived thermal plumes issuing from the outer boundaries, corresponding to classical Rayleigh–Bénard convection³⁴, Fig. 1b) and a large-scale thermochemical mode. The former was always fully developed in at least one of the layers before the onset of the latter. Here I focus on this new mode. Two regimes were observed (Fig. 2), depending on the ratio R_ρ of the chemical density anomaly $\Delta\rho$ to the thermal density anomaly^{35,36}:

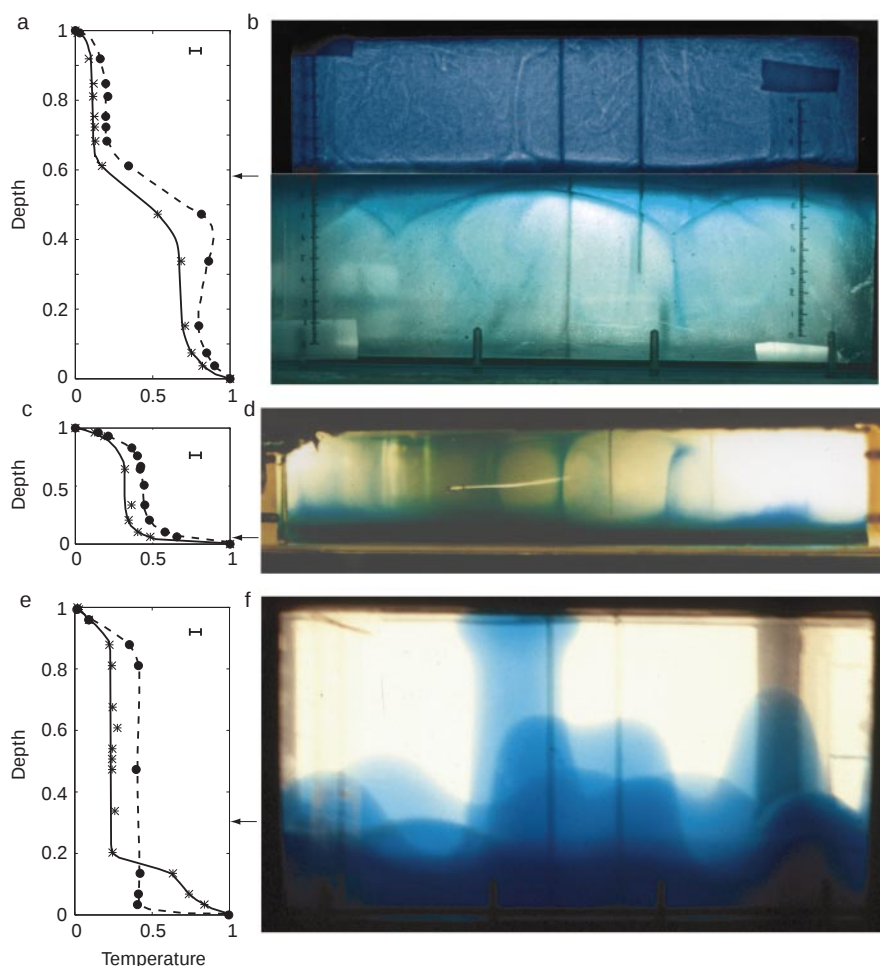


Figure 1 Snapshots and vertical temperature structure for the different regimes of convection. Time t is non-dimensional; it indicates the run time of the experiment divided by the thermal diffusion time. Temperature is non-dimensionalized by the temperature difference imposed across the tank, and depth by the tank thickness. The arrows mark the interface position at the beginning of each experiment. **a, b**, Strong stratification ($t = 0.093$; $R_\rho = 2.4$; $\gamma = 1.1 \times 10^{-2}$; $Ra_{lower} = 2.2 \times 10^5$; $Ra_{upper} = 8 \times 10^6$; $d_2/d_1 = 0.9$; $32 \times 32 \times 14.8$ cm tank); see text for definition of symbols. Panel **b** is a composite snapshot showing a direct view of the entrainment in the bottom, more viscous, layer and a shadowgraph of the top layer. The interface remains flat and sharp. Two scales of motions are visible: a large-scale circulation generated at the interface, and a small-scale thermal mode with plumes coming from the two outer boundaries. Entrainment only occurs through the interfacial mode: in the lower layer, this consists of the blue cusps and the thin sheets of blue, light, low-viscosity fluid (one can distinguish three cells); in the upper layer, it consists of the very thin conduit centres on top of a cell, which is not distorted until it encounters the upper cold thermal boundary layer. **a**, Vertical thermal structure associated with the cold descending currents in the cusps

(asterisks) and with the hot narrow plume conduit (filled circles). **c, d**, Entrainment through the thin tubular plumes when the bottom layer is very thin ($d_2/d_1 = 7$). Note the topography of the bottom layer. ($t = 0.0948$; $R_\rho = 0.45$; $\gamma = 2 \times 10^{-2}$; $Ra_{lower} = 40$; $Ra_{upper} = 3 \times 10^6$; $d_2/d_1 = 3$; $32 \times 32 \times 6$ cm tank). Rising thermochemical plumes are again associated with hot temperature profiles (filled circles) compared to the rest of the tank (asterisks). **e, f**, Weak stratification ($t = 0.0117$; $R_\rho = 0.2$; $\gamma = 8 \times 10^{-3}$; $Ra_{lower} = 9 \times 10^4$; $Ra_{upper} = 9 \times 10^6$; $d_2/d_1 = 3$; $32 \times 32 \times 14.8$ cm tank). Domes are forming and one of them has reached the upper boundary. It corresponds to the hot vertical profile (filled circles) which shows a homogeneous temperature over almost the whole depth on the tank, while other spots outside this dome show a stratified thermal structure (asterisks). During this experiment, 5 cycles were observed with a dimensionless periodicity around 0.0134. In **d** and **f**, the small-scale thermal mode was also present in the light fluid, but is not visible with direct light; the magnitude of its temperature fluctuations is indicated by the small horizontal bar in **a, c, e**. The vertical dark lines are the thermocouple probes.

$$R_p = \frac{\Delta\rho}{\rho\alpha\Delta T} \quad (2)$$

‘Stratified’ regime. When $R_p > 1$, thermal convection develops in two superimposed layers³² (Fig. 1b), separated by a thermal boundary layer at a relatively undeformed interface (Fig. 1a). The entrainment is asymmetric, with narrow cylindrical plumes in layer 2 (higher R_a) and two-dimensional sheets in layer 1 (Fig. 1b). These thermochemical features involve temperature fluctuations much bigger than those of purely thermal convection³⁴: because the layers are miscible, any instability starting from one side of the interface must entrain, by viscous coupling, a film of more stable material from the other side; thus for instabilities to occur at the interface, their temperature must increase to overcome the stabilizing effect of the chemical heterogeneities that they entrain. If one layer is too thin to convect—as may be the case for the D'' layer at the base of the mantle—a conductive thermal gradient develops across it. The system then has only two thermal boundary layers, one of them (the lower, say) being thicker as it is stratified in density^{25,37,38} (Fig. 1c). Hot cylindrical plumes rising from it (Fig. 1d) are associated with heat flux anomalies of up to a factor of three relative to the ambient flux. These ‘hotspots’ form a stable but slowly migrating polygonal

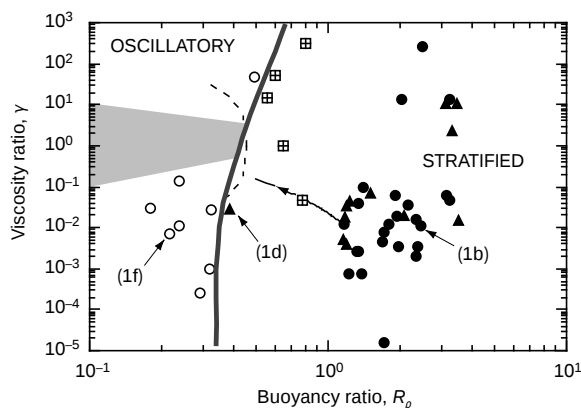


Figure 2 Different convective regimes as a function of the buoyancy ratio R_p and the viscosity ratio γ . A thin hot D'' layer would be in the region $\gamma > 10^3$, and a stratified mantle with a viscosity increase with depth would be in the region $\gamma < 10^{-1}$. The thick line separates the oscillatory ‘doming’ mode from the strictly stratified regime. The dashed line represents the results of the marginal stability analysis. In grey is the domain where domes form but break up and mix within one cycle. Each symbol represents an experiment: open circles, domes involving the whole tank; filled circles, two-layer convection with hotspot formation, the line issuing from one of those dots shows the typical time evolution of an experiment; triangles, one layer too thin to convect and hotspot formation; patterned squares, hybrid regime with high interfacial topography and hotspots. The three experiments shown in Fig. 1 are indicated: (1b), (1d), (1f).

pattern^{34,38}. The effect of a strong stratification is thus to focus heat and stabilize features compared to classical thermal convection.

‘Doming’ regime. When $R_p < 0.35$ – 0.55 , the interface deforms in large domes (Fig. 1f) which move up and down quasi-periodically. Doming is possible because the density at the heated bottom of the lower layer is smaller than the density in the cooled top of the upper layer, despite the stabilizing jump across the interface. The direction of doming is determined by the balance between tendencies to penetrate the less viscous layer and the thicker layer. Thus cold, less-viscous domes sometimes invade the more viscous layer when the latter is thicker. More relevant to the mantle are hot rising domes, which form when the lower layer is heated (by conduction or convection) to the point where its thermal buoyancy counterbalances the stable chemical stratification and the viscous forces (Fig. 1e). When such a dome reaches the top of the tank, it cools and loses its thermal buoyancy. Being compositionally denser, it falls back to the bottom. At high R_a , it may even detach from its base before sinking again. When the interlayer viscosity ratio γ is close to unity, mixing occurs before the dome descends and the oscillation ceases³⁹. The number of cycles observed depends strongly on the viscosity and depth ratios, and more weakly on the Rayleigh number; I observed at most 5 cycles before complete mixing. After one cycle, the domes may contain encapsulated ‘blobs’ of less viscous and dense fluid. For $\gamma > 10$, hot thermochemical plumes typically form on the domes (Fig. 3).

Hotspots and superswells on Earth

Although the experimental system is greatly simplified (no plate tectonics, no radioactive heating, no temperature- or pressure-dependence of ν , κ , α), it gives important clues to the behaviour of different regions of the mantle. In particular, Fig. 2 allows us to estimate whether the mantle today is wholly in the stratified regime, wholly in the doming regime, or in different regimes at different lateral positions. Because the values of some of the relevant physical properties of the mantle (especially the viscosity and its variation as a function of temperature, depth and composition⁴⁰) are poorly known, I use ranges of values ($R_a = 10^5$ – 5×10^7 ; $\gamma = 10^{-3}$ – 10^3 ; $\Delta\rho/\rho = 0.1$ – 5% ; $d_2/d_1 = 0.2$ – 40) in the calculations below.

From equation (2), the critical density contrast separating the stratified and doming regimes is $\Delta\rho_{\text{crit}}/\rho = \alpha\Delta T(R_p)_{\text{crit}}$, where $(R_p)_{\text{crit}}$ is the critical buoyancy ratio. Figure 4 shows $\Delta\rho_{\text{crit}}/\rho$ as a function of ΔT for $(R_p)_{\text{crit}} = 0.35$ and 0.55 , and $\alpha = 5 \times 10^{-5} \text{ K}^{-1}$ (typical for the uppermost mantle), and $1 \times 10^{-5} \text{ K}^{-1}$ (deep mantle)⁴¹. The transition between two-layer convection and doming occurs when $\Delta\rho_{\text{crit}}/\rho$ is of the order of 1%. Numerical experiments^{38,42} suggest that doming may occur for even higher values of $\Delta\rho/\rho$ when internal heating is concentrated in mantle chemical heterogeneities⁵. But, because $\Delta\rho_{\text{crit}}/\rho$ depends linearly on α , and α decreases with depth⁴¹, $\Delta\rho_{\text{crit}}/\rho$ will also decrease with depth. Thus a given density contrast $\Delta\rho/\rho$ can generate doming if

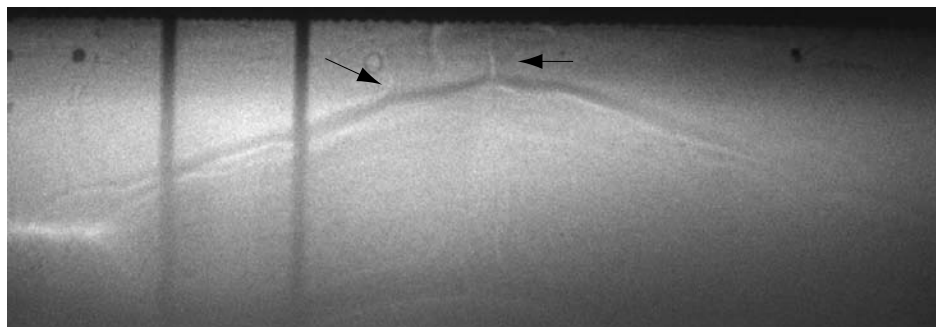


Figure 3 Doming regime. Shadowgraph (top half of the tank) showing a close-up view of a rising dome. Note the plume (constituted of a thin conduit and a big cap) on the top of the

dome. The remnant of another plume conduit can also be faintly seen. ($t = 0.203$; $R_p = 0.4$; $\gamma = 8 \times 10^{-3}$; $R_{a,\text{lower}} = 9 \times 10^5$; $R_{a,\text{upper}} = 8 \times 10^6$; $d_2/d_1 = 1$).

the interface is in the mid-mantle, or generate stratified convection if it is near the bottom of the mantle (Fig. 4, point A, marked by a plus sign). This suggests a unified physical explanation for three important features of convection in the present-day mantle: super-swells, 'weak' hotspots, and 'cold' hotspots.

Superswells. Consider an interface in the mid-mantle. The appropriate value of α at this depth is about $(2-3) \times 10^{-5} \text{ K}^{-1}$. For a typical temperature increase across the mantle of about 2,500 K, Fig. 4 then shows that stratified convection would require $\Delta\rho/\rho > 2\%$ (refs 29, 32), which seems ruled out today⁴. A mid-mantle interface will therefore give rise to doming (Fig. 1f), with domes 1,000–6,000 km in lateral extent oscillating up and down on a timescale of 100 Myr to 1 Gyr, and crowned by several small hot plumes (Fig. 3). The two present-day superplumes may therefore represent two different stages of the same phenomenon: a thermochemical blob ascending and descending in the mantle. The African superswell would correspond to the early stage of dome formation and ascent, while the Polynesian superswell would correspond to a more mature dome, which had already gone through a whole cycle^{15,22}.

'Weak' hotspots. Equation (2) shows that the stratification parameter R_p is proportional to an effective temperature difference ΔT . Now the temperature difference across the interface between the two layers (Fig. 4, point B, marked by a plus sign) is less than that over the whole mantle (Fig. 4, point A). Therefore, the 'local' R_p at the upper surface of a dome will be higher than the 'global' R_p for the whole mantle. Because the stratification at the top of a dome then is locally 'strong', thin hot plumes entraining dome material are expected. I suggest that such plumes may produce the Earth's 'weak' hotspots²⁴.

'Cold' hot spots. Consider now an interface within the core–mantle boundary layer. Because the temperature jump there exceeds 1,000 K (ref. 43) and $\alpha \approx 10^{-5} \text{ K}^{-1}$, strong stratification occurs for $\Delta\rho/\rho > 0.4\%$ (Fig. 4). A thin, chemically dense sublayer at the base of the core–mantle boundary longer with $\Delta\rho/\rho$ of several per cent (ref. 6) would thus remain stable, and plumes would arise only from the portion of the thermal boundary layer above the sublayer (Fig. 1d). The present laboratory experiments and previous numerical studies^{25,37,38} show that the temperature drop across this portion of the thermal boundary layer is 20–70% of that across the whole layer (Fig. 1c). Scaled to the mantle, the experiments suggest plumes with thermal anomalies between 150 and 500 °C, diameters around

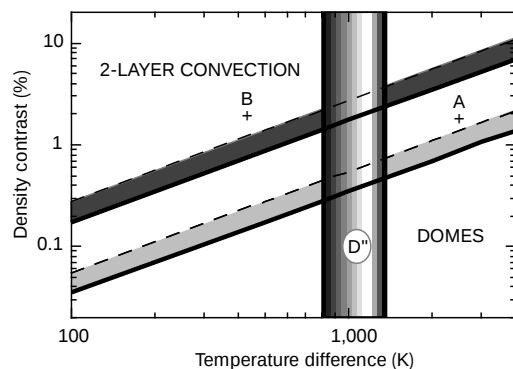


Figure 4 Critical density contrast $\Delta\rho/\rho$ calculated from equation (2) as a function of the temperature difference ΔT driving convection. Data are shown for $\alpha = 5 \times 10^{-5} \text{ K}^{-1}$ (typical of the uppermost mantle, diagonal dark-grey area) and $\alpha = 10^{-5} \text{ K}^{-1}$ (typical of the lowermost mantle, diagonal light-grey area). The solid line under each diagonal grey area has been calculated using $R_p = 0.35$, and the dashed line (above each diagonal grey area) using $R_p = 0.55$. The vertical column shows the local temperature range across the core–mantle boundary layer^{6,43}. D'' should be stratified and generates hotspots for heterogeneities greater than 0.4%. Point A is in the area where an interface in the lowermost mantle would stay horizontal, while an interface in the mid-mantle would dome. Point B characterizes the local stratification at one of these domes' interface.

100 km, and a characteristic spacing of 800–2,000 km, which survive until the total entrainment of the heterogeneity. These plumes may produce the Earth's 'cold' hotspots^{24,25}.

Temporal evolution and geochemical reservoirs

So far I have presented only 'snapshots' of thermochemical convection. However, such a system is intrinsically time-dependent. In both the stratified and the doming regimes, heterogeneities are entrained mechanically across the interface and eventually get completely homogenized. Some experiments even started in the stratified regime, and ended up in the doming regime (Fig. 2). If the mantle evolves similarly, the experiments reported here may provide an explanation for the long-term survival of geochemically distinct mantle reservoirs^{5,44}.

Doming. According to the experimental results, at most 5 doming cycles of duration between 100 Myr and 1 Gyr (depending on the viscosities and sizes of the reservoirs) are possible before complete mixing of the mantle. This implies that the present-day mantle regime is not older than 500 Myr to 5 Gyr: so doming is likely to be younger than the Earth. The preservation of a 'primitive' submantle throughout the age of the Earth thus requires mantle dynamics to have evolved to the present-day doming regime from a more strongly stratified initial stage.

Strong stratification. The time required to homogenize an initially stratified mantle can be estimated using a scaling law for the mixing rate. A simple balance of buoyancy and viscous forces^{34,45} implies that the volumetric flux of entrained material across the interface is:

$$Q_m = C_1 \rho \kappa L R_p^{-2} \text{Ra}^{1/3} \frac{1}{1 + (\gamma R_p)^{-1}} \quad (3)$$

where C_1 is an experimental constant and L is the lateral distance between plumes. Equation (3) reflects the fact that it is more difficult to entrain a viscous and/or strongly stratified fluid than an inviscid and/or weakly stratified one. Figure 5 shows that equation (3) agrees well with the experimental results when $C_1 = (6.1 \pm 0.5) \times 10^{-3}$. With this value, equation (3) implies that a mantle plume would entrain chemically dense material at a rate between 0.1% and 15% of its total volume flux, and that it would take between 0.3 Gyr (for a thin core–mantle boundary layer) and 40 Gyr (for strong stratification at mid-mantle depth) to erase the heterogeneities spread on an area of size L^2 beneath it. These times are lower bounds because the experiments do not include the effect of continuous reinjection of heterogeneous material into the mantle by subduction. Long-lived geochemical reservoirs are therefore to be expected in a chemically heterogeneous convecting mantle.

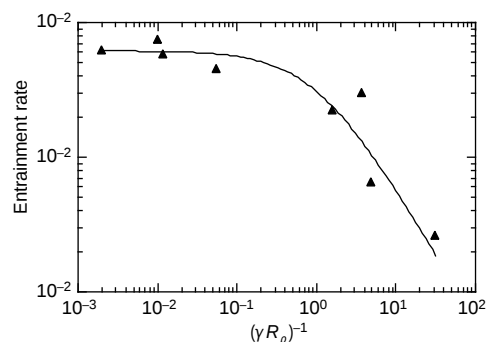


Figure 5 Volumetric entrainment rate of a cylindrical plume as a function of $(\gamma R_p)^{-1}$. The entrainment rate shown is non-dimensional; it represents the experimental rate divided by the convective scale $Q = \rho \kappa L R_p^{-2} \text{Ra}^{1/3}$. The triangles represent the experimental measurements for the experiments with a layer too thin to convect, while the solid line stands for the law $1/(1 + (\gamma R_p)^{-1})$ coming from equation (3) in the text.

Discussion

Chemical density heterogeneities of 0.1–2% have far from negligible effects on mantle dynamics. They produce hot domes and plumes which provide a physical explanation for the quasi-periodic superswells and relatively fixed hotspots seen on Earth. Moreover, a doming regime cannot have persisted for the whole age of the Earth. We thus have to abandon the notion of a single style of mantle dynamics through time in favour of an evolution from an early, strongly stratified regime to the present-day superswell and hotspot regime. Thermochemical convection can then account for both present-day patterns (distribution of hotspots and superswells, seismic tomography) and the existence of long-lived reservoirs inferred from geochemical data.

Thermochemical convection is also likely to have a strong influence on the dynamics of the core. Domes and hotspots generate high-amplitude heat flux anomalies of up to four times the ambient flux, which will result in a strongly inhomogeneous and time-dependent thermal boundary condition at the core–mantle boundary^{46,47}. This spatiotemporal variability may be related to the recurrence pattern of magnetic quiescence^{14,22}.

Thermochemical convection poses further challenging questions that remain to be answered. Will seismology in the future be able to detect small chemical density heterogeneities and non-horizontal interfaces? How dependent are geochemical inferences on the time-history of the reservoirs? How do additional factors such as plate tectonics, phase transitions, and the variable properties of mantle materials affect the results presented here? Further experimental work is in progress to characterize fully the doming regime, and to quantify the effect of temperature-dependent viscosity. □

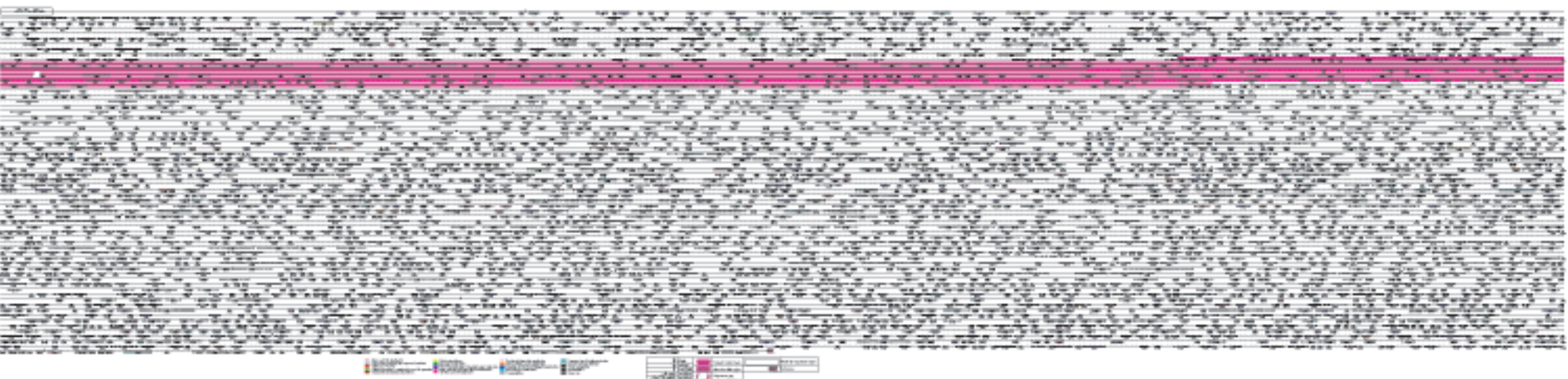
Received 6 April; accepted 2 November 1999.

1. Olson, P., Silver, P. G. & Carlson, R. W. The large scale structure of convection in the Earth's mantle. *Nature* **344**, 209–215 (1990).
2. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
3. Jackson, I. Elasticity, composition and temperature of the Earth's lower mantle: a reappraisal. *Geophys. J. Int.* **134**, 291–311 (1998).
4. Bina, C. R. in *Ultra-high Pressure Mineralogy* (ed. Hemley, R. J.) 205–239 (Reviews in Mineralogy Vol. 37, Mineralogical Society of America, Washington, DC, 1998).
5. Zindler, A. & Hart, S. Chemical geodynamics. *Annu. Rev. Earth Planet. Sci.* **14**, 493–571 (1986).
6. Lay, T., Williams, Q. & Garnero, E. J. The core–mantle boundary layer and deep Earth dynamics. *Nature* **392**, 461–467 (1998).
7. Van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
8. Forte, A. M. & Woodward, R. L. Seismic-geodynamics constraints on three-dimensional structure, vertical flow, and heat transfer in the mantle. *J. Geophys. Res.* **102**, 17981–17994 (1997).
9. Wilson, J. T. Evidence from islands on the spreading of the ocean floor. *Can. J. Phys.* **41**, 863–868 (1963).
10. Morgan, W. J. Plate motions and deep mantle convection. *Nature* **230**, 42–43 (1971).
11. Richards, M. A., Duncan, R. A. & Courtillot, V. E. Flood basalts and hot-spot tracks: Plume heads and tails. *Science* **246**, 103–107 (1989).
12. Sleep, N. H. Hotspots and mantle plumes: some phenomenology. *J. Geophys. Res.* **95**, 6715–6736 (1990).
13. Loper, D. E. & Stacey, F. D. Mantle plumes and the periodicity of magnetic field reversals. *Geophys. Res. Lett.* **13**, 1525–1528 (1986).
14. Larson, R. L. & Olson, P. Mantle plumes control magnetic reversal frequency. *Earth Planet. Sci. Lett.* **107**, 437–447 (1991).
15. McNutt, M. K. Superswells. *Rev. Geophys.* **36**, 211–244 (1998).
16. Nyblade, A. A. & Robinson, S. W. The African superswell. *Geophys. Res. Lett.* **21**, 765–768 (1994).
17. McNutt, M. K. & Fisher, K. M. in *Seamounts, Islands and Atolls* (eds Keating, B. H. et al.) 25–34 (Geophys. Monogr. Ser. 43, American Geophysical Union, Washington DC, 1987).
18. Masters, G., Johnson, S., Laske, G. & Bolton, H. A shear-velocity model of the mantle. *Phil. Trans. R. Soc. Lond. A* **354**, 1385–1410 (1996).
19. Hart, S. R. A large-scale isotope anomaly in the Southern Hemisphere mantle. *Nature* **309**, 753–757 (1984).
20. Cazenave, A. & Thoraval, C. Mantle dynamics constrained by degree 6 surface topography, seismic tomography and geoid: Inference on the origin of the South Pacific Superswell. *Earth Planet. Sci. Lett.* **122**, 207–219 (1994).
21. Vinnik, L., Chevrot, S. & Montagner, J.-P. Evidence for a stagnant plume in the transition zone. *Geophys. Res. Lett.* **24**, 1007–1010 (1997).
22. Larson, R. L. Geological consequences of superplumes. *Geology* **19**, 547–550 (1991).
23. Bercowski, D., Schubert, G. & Glatzmaier, G. A. Three-dimensional spherical models of convection in the Earth's mantle. *Science* **244**, 950–955 (1989).
24. Albers, M. & Christensen, U. R. The excess temperature of plumes rising from the core–mantle boundary. *Geophys. Res. Lett.* **23**, 3567–3570 (1996).
25. Farnetani, C. G. Excess temperature of mantle plumes: the role of chemical stratification across D". *Geophys. Res. Lett.* **24**, 1583–1586 (1997).
26. Hoffman, N. R. A. & McKenzie, D. P. The destruction of geochemical heterogeneities by differential motions during mantle convection. *Geophys. J. R. Astron. Soc.* **82**, 163–206 (1985).
27. Christensen, U. Mixing by time-dependent convection. *Earth Planet. Sci. Lett.* **95**, 382–394 (1989).
28. Kellogg, L. H. Chaotic mixing in the Earth's mantle. *Adv. Geophys.* **34**, 1–33 (1993).
29. Gurnis, M. & Davies, G. F. The effect of depth-dependent viscosity on convective mixing in the mantle and the possible survival of primitive mantle. *Geophys. Res. Lett.* **13**, 541–544 (1986).
30. Spence, D. A., Ockendon, J. R., Wilmott, P., Turcotte, D. L. & Kellogg, L. H. Convective mixing in the mantle: the role of viscosity differences. *Geophys. J.* **95**, 79–86 (1988).
31. Manga, M. Mixing of heterogeneities in the mantle—Effects of viscosity differences. *Geophys. Res. Lett.* **23**, 403–406 (1996).
32. Olson, P. An experimental approach to thermal convection in a two-layered mantle. *J. Geophys. Res.* **89**, 11293–11301 (1984).
33. van Keken, P. E. & Ballentine, C. J. Whole-mantle versus layered mantle convection and the role of a high-viscosity lower mantle in terrestrial volatile evolution. *Earth Planet. Sci. Lett.* **156**, 19–32 (1998).
34. Davaille, A. Two-layer thermal convection in miscible viscous fluids. *J. Fluid Mech.* **379**, 223–253 (1999).
35. Richter, F. M. & Johnson, C. E. Stability of a chemically layered mantle. *J. Geophys. Res.* **79**, 1635–1639 (1974).
36. Richter, F. M. & McKenzie, D. P. On some consequences and possible causes of layered convection. *J. Geophys. Res.* **86**, 6133–6124 (1981).
37. Christensen, U. Instability in a hot boundary layer and initiation of thermochemical plumes. *Ann. Geophys.* **2**, 311–320 (1984).
38. Tackley, P. J. in *The Core–Mantle Boundary Region* (eds Gwinn, M. et al.) 231–255 (AGU Monogr. American Geophysical Union, Washington DC, 1998).
39. Olson, P. & Kincaid, C. Experiments on the interaction of thermal convection and compositional layering at the base of the mantle. *J. Geophys. Res.* **96**, 4347–4354 (1991).
40. Karato, S. & Wu, P. Rheology of the upper mantle: a synthesis. *Science* **260**, 771–778 (1993).
41. Chopelas, A. & Boehler, R. Thermal expansivity in the lower mantle. *Geophys. Res. Lett.* **19**, 1983–1986 (1992).
42. Kellogg, L. H., Hager, B. H. & van der Hilst, R. D. Compositional stratification in the deep mantle. *Science* **283**, 1881–1884 (1999).
43. Boehler, R. Temperatures in the Earth's core from melting point measurements of iron at high static pressures. *Nature* **363**, 534–536 (1993).
44. Allègre, C. J. & Lewin, E. Isotopic systems and stirring times of the earth's mantle. *Earth Planet. Sci. Lett.* **136**, 629–646 (1995).
45. Sleep, N. H. Gradual entrainment of a chemical layer at the base of the mantle by overlying convection. *Geophys. J.* **95**, 437–447 (1988).
46. Montague, N. L., Kellogg, L. H. & Manga, M. High-Rayleigh number thermochemical models of a dense boundary layer in D". *Geophys. Res. Lett.* **25**, 2345–2348 (1998).
47. Girard, F. & Davaille, A. Dynamics of an heterogeneous layer at the base of the mantle: an experimental approach. *Eos* **79**, 617 (1998).

Acknowledgements

I thank C. Allègre, C. Jaupart, K. Turekian, P. Molnar, N. Ribe and G. Veronis for discussions; G. Bienfait, F. Girard, A. Lee, W. Phelps and W. Sacco for help in the laboratory; and M. McNutt, M. Manga and P. Tackley for comments on the manuscript. Part of this work was done at the Department of Geology and Geophysics at Yale University, USA.

Correspondence and requests for materials should be addressed to the author (e-mail: davaille@ipgp.jussieu.fr).



A structural change in the kinesin motor protein that drives motility

Sarah Rice*, Abel W. Lin†, Daniel Safer‡, Cynthia L. Hart§, Nariman Naber||, Bridget O. Carragher¶, Shane M. Cain†, Elena Pechatnikova#, Elizabeth M. Wilson-Kubalek†, Michael Whittaker†, Edward Pate*, Roger Cooke||, Edwin W. Taylor#, Ronald A. Milligan† & Ronald D. Vale*§

§ The Howard Hughes Medical Institute, and the * Departments of Cellular and Molecular Pharmacology and || Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

† Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92037, USA

‡ Department of Physiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

¶ Beckman Institute, Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois 61801, USA

Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60637, USA

* Department of Pure and Applied Mathematics, Washington State University, Pullman, Washington 99164, USA

Kinesin motors power many motile processes by converting ATP energy into unidirectional motion along microtubules. The force-generating and enzymatic properties of conventional kinesin have been extensively studied; however, the structural basis of movement is unknown. Here we have detected and visualized a large conformational change of a ~15-amino-acid region (the neck linker) in kinesin using electron paramagnetic resonance, fluorescence resonance energy transfer, pre-steady state kinetics and cryo-electron microscopy. This region becomes immobilized and extended towards the microtubule 'plus' end when kinesin binds microtubules and ATP, and reverts to a more mobile conformation when γ -phosphate is released after nucleotide hydrolysis. This conformational change explains both the direction of kinesin motion and processive movement by the kinesin dimer.

Microtubule-based motors of the kinesin superfamily are involved in intracellular transport, mitosis and meiosis, control of microtubule dynamics and signal transduction pathways¹. To execute these various activities, different kinesin family members have distinct protein domains that confer unique cargo binding, regulatory and oligomerization properties. However, all kinesins have a homologous ~330-amino-acid 'catalytic core' that binds ATP and microtubules². This catalytic core bears a striking structural similarity to the nucleotide-binding core of the actin-based motor myosin, suggesting that these two molecular motors originated from a common ancestor^{3,4}.

Conventional kinesin, a dimer of identical ~120K (120,000 relative molecular mass (M_r)) heavy chains that transports organelles towards the microtubule plus end, has served as an important model system for understanding biological motility. Conventional kinesin (hereafter called 'kinesin') is a highly processive motor that can take more than 100 consecutive 8-nm steps (the distance between two α/β subunits of tubulin) before dissociating^{5,6}. Such processive motion requires coordination between the two motor domains in the kinesin dimer⁷⁻⁹, and an alternating-site enzymatic mechanism has been proposed¹⁰⁻¹². Although not processive, kinesin monomers produced by truncation also generate motility, provided that several monomers interact simultaneously with a microtubule⁷⁻⁹.

Myosin-based movement is believed to be generated by the angular rotation of a long α -helix and its associated light chains that occurs after nucleotide hydrolysis and phosphate release¹³⁻¹⁶. However, kinesin lacks an elongated lever arm and thus may use another type of mechanical element. The 'neck linker', a ~15-amino-acid segment carboxy-terminal to the catalytic core (Fig. 1), has been suggested to be important for kinesin force generation, as its sequence is highly conserved among plus-end- but not minus-end-directed kinesin motors^{2,17}. In support of this idea, analyses of protein chimaeras between opposite polarity motors in the kinesin superfamily implicate the neck region as a governor of directionality¹⁷⁻¹⁹, and mutagenesis of the neck linker severely impairs motility while having relatively little effect on ATPase activity (R. Case and R.D.V., manuscript submitted). Cryo-electron microscopy

(cryo-EM) studies have also found nucleotide-dependent conformational changes in the microtubule-bound kinesin dimer²⁰⁻²², which could potentially involve the neck linker. Here we have directly measured conformations of the kinesin neck linker in different nucleotide- and microtubule-binding states, and the results have allowed us to formulate a structural model for how kinesin moves along a microtubule.

Generation of a cysteine-light kinesin

To measure structural changes, we engineered a human kinesin without solvent-exposed cysteines, which has normal enzymatic and motile properties (see Methods). We introduced cysteines at defined locations for the attachment of biophysical probes. Cysteine 333 was chosen as a primary site, because its location is ideal for detecting potential motions of the neck linker relative to the catalytic core. Modification of this cysteine with probes had little or no effect on motility (see Methods). Results were confirmed by also attaching probes at C328 and, in some instances, C330. To simplify the interpretation of results, a kinesin monomer (K349) was used, because the two heads of a microtubule-bound kinesin dimer are thought to be in different conformational states¹⁰⁻¹².

Mobility changes detected by EPR

We measured electron paramagnetic resonance (EPR) spectra derived from a paramagnetic probe attached to C333 (Fig. 2). We present our results in terms of a cycle in which soluble kinesin docks onto a microtubule, binds ATP, hydrolyses it and releases products.

In solution, the EPR spectra were superimposable in several nucleotide states (ADP, AMP-PNP, ADP- AlF_4^- and ADP- BeF_x ; last two not shown). The shape of the three sharp lines indicated that the probe was highly mobile, and behaved as if attached to a mobile protein loop²³. When kinesin bound to the microtubule in the nucleotide-free state, the EPR spectrum was comparable to kinesin in solution; however, when AMP-PNP (a very slowly hydrolysed ATP analogue), ADP- AlF_4^- or ADP- BeF_x (data not shown) bound to microtubule-attached kinesin, a new spectral component appeared with peaks at high and low fields (Fig. 2). These spectra reveal a substantial decrease in mobility, which is indicative of a probe

attached to a rigid secondary structure element²³. In the microtubule-bound ADP state, the EPR probe returned to a state of high mobility. Probes introduced at residues 328 and 330 produced results similar to those described above (data not shown), indicating that the spectral changes did not arise simply from altered local probe–protein interactions, but rather from a conformational change involving most of the neck linker.

Analysis of the temperature dependence of the spectra between 4°C and 35°C using the van't Hoff equation²⁴ showed that the change in enthalpy (ΔH) for the transition to the AMP-PNP-induced conformation was -20 , -15 and -8 kcal mol⁻¹ for residues 333, 330 and 328, respectively. Again, these large values indicate a significant protein conformational change, rather than an altered local probe–protein interaction; in addition, the variation of ΔH with position on the neck linker is consistent with a sequential zippering of the neck linker beginning at its N-terminal connection to the catalytic core. Collectively, our results indicate that the neck linker is mobile and becomes more ordered when the microtubule-bound motor binds an ATP analogue. After γ -phosphate release, the neck linker returns to its mobile conformation.

Distance changes detected by FRET

To detect neck linker movements in different nucleotide states, fluorescence resonance energy transfer (FRET) was measured between dyes introduced at C333 and a nearby residue (C220) on the catalytic core (Fig. 1; Table 1). A nucleotide-dependent conformational change was not observed in solution, as the energy transfer efficiencies were similar with ADP or AMP-PNP ($\sim 85\%$). For kinesin bound to microtubules in the nucleotide-free state, the energy transfer efficiency was lower than in solution (76%). A significant increase in energy transfer (93%) was observed when microtubule-docked kinesin bound AMP-PNP, indicating that the two probes moved closer together by ~ 10 Å. We also measured FRET between residues 328 and 220, and obtained similar results to

those described above (data not shown). Most of the AMP-PNP-induced motion must come from residue 333, as cryo-EM results (see below) indicate that residue 220 moves by less than 5 Å between the nucleotide-free and AMP-PNP states. The calculated distances between fluorescent probes (Table 1) and cryo-EM studies (see below) both suggest that the neck linker position most closely resembles the rat kinesin crystal structure in the microtubule-bound AMP-PNP state.

To probe the conformation of the neck linker without modifying its residues, we fused a green fluorescent protein (F64L, S65T GFP) to the C terminus of the neck linker at residue 339 (Fig. 1) and used it as a fluorescence donor for energy transfer to tetramethylrhodamine at C220. For microtubule-bound kinesin, FRET was higher in the AMP-PNP than in the nucleotide-free state (Fig. 3). Thus, GFP at the end of the neck linker moves closer to the tip of the catalytic core upon AMP-PNP binding, in agreement with the measurements described above.

We also examined FRET between C333 and C20, which is located at the opposite end of the catalytic core from 220 (Fig. 1). The change in FRET efficiency between nucleotide-free and AMP-PNP states on the microtubule ($64.1 \pm 6.3\%$ and $42.3 \pm 5.3\%$ energy transfer, respectively) was opposite to that observed between C333 or K339–GFP and C220. Thus, in the nucleotide-free state, FRET reveals that the neck linker is positioned away from residue 220 and towards residue 20, which is most easily explained if it detaches from the catalytic core. In contrast, FRET between C20 and another catalytic core residue (C188) did not change significantly between nucleotide-free and AMP-PNP states on the microtubule ($88.3 \pm 3\%$ and $87.0 \pm 1.5\%$ energy transfer, respectively).

In conclusion, FRET shows that the neck linker undergoes a large motion directed towards the tip of the catalytic core when microtubule-bound kinesin binds an ATP analogue. In contrast, for kinesin in solution, the FRET signal was nucleotide-independent and intermediate in value between the microtubule-bound

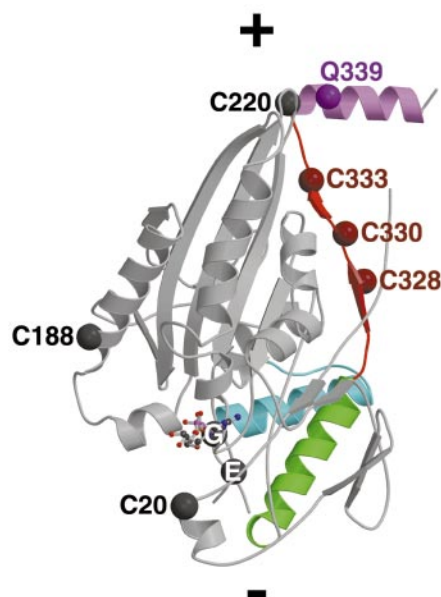


Figure 1 Structure of the rat kinesin monomer²⁵ (numbered as in human kinesin). The kinesin structure is oriented as it docks onto the microtubule in Fig. 4 (+ and – indicate the direction of the microtubule plus and minus ends). The neck linker (residues 323–335) is red, the neck coiled-coil (336–349) is purple, $\alpha 4$ helix (256–270) and L12 (the microtubule binding loop; 270–281) are blue, and the $\alpha 6$ helix (residues 309–320) is green. ADP is shown in the active site as a ball-and-stick model. The mutated residues G234 and E236 near the nucleotide site are indicated by circles labelled 'G' and 'E'. Residue 339 marks the junction point in the K339–GFP fusion protein. Cysteines were introduced at positions 20, 188, 220, 328 and 333 (circles at C α positions) for site-specific labelling in a 'cys-light' kinesin.

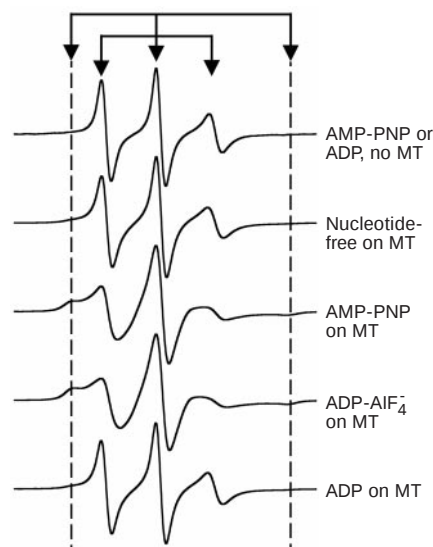


Figure 2 Electron paramagnetic resonance spectra for kinesin C333–MSL in several nucleotide states, both free in solution and bound to microtubules. The two sets of resonance peaks are indicated by arrows. The narrower resonance peaks are indicative of a highly mobile probe, whereas the wider set of resonance peaks (highlighted by the vertical dashed lines) are indicative of restricted mobility and emerge in the triphosphate states on microtubules. This mobility shift can be modelled by a restriction in probe motion from a cone angle of $\sim 120^\circ$ to 32° (ref. 50).

nucleotide-free and AMP-PNP states, suggesting that the neck linker may exist in docked and mobile conformations. In support of this idea, solution scattering studies have shown that the kinesin dimer can adopt multiple configurations^{27,28}. Furthermore, the neck linker is docked and disordered in the ADP crystal structures of rat^{25,26} and human³ kinesins, respectively. This suggests that the crystallization conditions for rat kinesin may have shifted a conformational equilibrium of the neck linker towards docking onto the catalytic core in its ADP state. Precedence for this exists in myosin, for which two markedly different crystal conformations were obtained with bound ADP-BeF_x^{14,29}, and an ADP structure has been reported to mimic an ATP configuration³⁰.

Structural changes visualized by cryo-EM

We sought to visualize the position of the neck linker directly by cryo-EM (Fig. 4). As the 15-amino-acid neck linker is not visible in our electron image maps, we used a gold cluster attached to C333 as a reporter of its position. In the nucleotide-free state, two density peaks for the gold particle were observed: one lying towards the minus end (called the -30° position relative to the axis perpendicular to the microtubule), and the second lying towards the plus end ($+125^\circ$ position) (Fig. 4a). The occurrence of two peaks, when a single residue was labelled, indicates that the neck linker exists in an equilibrium between two positions that are similar in energy. Further support for this idea comes from the finding that the relative strengths of the $+125^\circ$ and -30° peaks vary among independent 3D maps of the nucleotide-free state (data not shown). The above results reveal that the neck linker can attach to the catalytic core in the nucleotide-free state, which would seem to contradict the high mobility measured in the EPR studies. However, EPR shows that the neck linker in nucleotide-free and ADP states becomes increasingly immobilized when cooled to 4°C . Thus, the weak binding sites observed by cryo-EM may be populated during the freezing process.

For the motor-microtubule complex with AMP-PNP, the gold density was located at a third position, which is also near the tip of the catalytic core ($+50^\circ$ position) (Fig. 4b). This location also corresponds roughly to the position of the second head in the kinesin dimer in the microtubule-bound, AMP-PNP state²². A similar position (within $5\text{ \AA}$) was observed for the ADP- AlF_4^- state (Fig. 4c). For both AMP-PNP and ADP- AlF_4^- , the occurrence of a single, strongly-diffracting peak suggested nearly 100% occupancy at this single position. Thus, after the motor binds AMP-PNP, the neck linker undergoes a marked transition from dual site occupancy at positions $+125^\circ$ and -30° to that of a single position at $+50^\circ$.

Table 1 FRET between donor (coumarin, CPM) and acceptor (tetramethylrhodamine, TMR) probes attached to the neck linker (C333) and the catalytic core (C220) in wild-type kinesin monomer and two ATP nonhydrolysing mutants

Kinesin (K349–C220, C333)	Nucleotide	MT	Energy transfer (%)
Wild type	ADP	–	87.6 ± 3.3
	AMP-PNP	–	84.4 ± 0.2
	Nucleotide-free	+	76.4 ± 2.8
	AMP-PNP	+	93.4 ± 1.3
G234A	ADP	–	70.9 ± 6.1
	Nucleotide-free	+	70.1 ± 3.5
	ATP	+	75.1 ± 6.8
	AMP-PNP	+	78.4 ± 5.2
E236A	ADP	–	83.1 ± 4.5
	Nucleotide-free	+	69.9 ± 7.3
	ATP	+	95.1 ± 1.3
	AMP-PNP	+	92.3 ± 5.4

The mean and standard deviation are shown for at least three measurements. The maximum limiting anisotropy of the kinesin-bound probes did not change in the nucleotide-free and AMP-PNP states on microtubules (0.27 for CPM and 0.23 for TMR), indicating that the increase in energy transfer efficiency is due to a change in distance rather than a probe re-orientation. For wild type kinesin, the estimated probe–probe distances in the nucleotide-free and AMP-PNP states were about $34\text{ \AA}$ and $24\text{ \AA}$, respectively. MT, microtubule; +, presence; –, absence of microtubules.

When the gold-labelled motor was bound to microtubules in the presence of 5 mM ADP, densities at the $+125^\circ$ and -30° positions were again observed (Fig. 4d). Similar results were obtained with a gold cluster attached to residue 328 (data not shown), indicating that most of the neck linker undergoes a similar nucleotide-dependent conformational change.

In contrast to the pronounced movements of the neck linker, the 3D maps of unlabelled K349 show only subtle differences between the AMP-PNP and nucleotide-free states (Fig. 4e). In addition, when the catalytic core was labelled at position 220, the gold density moved only slightly ($<5\text{ \AA}$) between the nucleotide-free and AMP-PNP states (Fig. 4e, f). Thus, although small conformational changes may occur, the catalytic core does not shift its orientation on the microtubule upon nucleotide binding.

The orientation of the atomic structure in the cryo-EM maps must be established to determine how the neck linker is positioned in the catalytic core in its different nucleotide states; however, three different docking models have been reported^{21,31–33}. To resolve this controversy, we labelled residues 20 and 188 simultaneously with gold particles (Fig. 4h). Residues 20, 188 and 220 form a triangle on the catalytic core, and the three gold positions allowed us to orient the crystal structure unambiguously in the electron microscopy maps (details will be reported elsewhere). Figure 1 shows the docking orientation of the kinesin crystal structure from the same view as Fig. 4. The results agree closely with our previous model³¹, but not with other reported docking orientations^{21,32}. In our docking of the catalytic core, the location of residue 333 in the kinesin crystal structure corresponds to the $+50^\circ$ gold position observed with ATP analogues (compare Figs 1 and 4). At atomic resolution, however, the microtubule-bound ATP state may not be identical to the crystal structure, which reveals relatively weak contacts between the neck linker and the catalytic core. It is possible that subtle rearrangements strengthen the neck linker–catalytic core interactions in the presence of microtubules and ATP analogues.

Thus, cryo-EM shows that the catalytic core maintains its overall position and orientation on the microtubule throughout the ATPase cycle. In contrast, the neck linker exhibits a striking

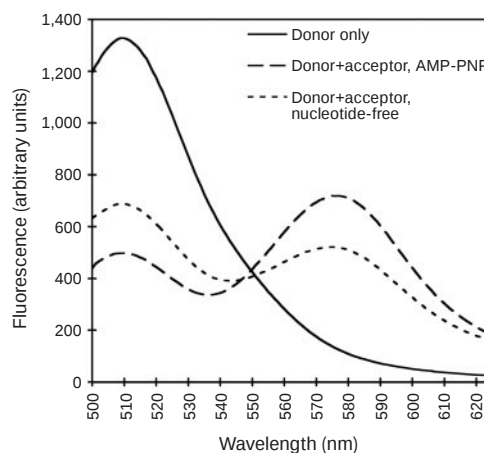


Figure 3 Nucleotide-dependent energy transfer between GFP (donor) at the end of the neck linker and tetramethylrhodamine (acceptor) at C220. The emission spectrum of microtubule-bound K339–GFP without acceptor is shown by the solid black line; the spectrum was identical in 5 mM Mg-AMP-PNP and in the absence of nucleotide on microtubules (MT). In the presence of TMR acceptor at position 220, GFP fluorescence is quenched. There is significantly greater donor quenching (488 nm) and a corresponding enhancement of acceptor fluorescence (575 nm) in the presence of AMP-PNP than in the absence of nucleotide (energy transfer efficiencies of $60.6 \pm 2.3\%$ and $52.1 \pm 5.1\%$, respectively). The difference in energy transfer was somewhat lower than with other fluorescent probes (Table 1), probably because GFP was somewhat mobile, even in the AMP-PNP state, as the result of its flexible connection to the kinesin neck linker.

nucleotide-dependent behaviour. In the absence of bound nucleotide, the neck linker is distributed between forward ($+125^\circ$) and backward (-30°) pointing states that are separated by ~ 50 Å. On the basis of these two positions, we predict that the neck linker is able to pivot about a point that lies near the C terminus of $\alpha 6$. Subsequent binding of ATP analogues to the catalytic core locks the neck linker into a second forward pointing position at $+50^\circ$. After phosphate release, however, the neck linker is again mobile and weakly interacts with two sites on the catalytic core.

ATPase kinetics of G234A and E236A mutants

We identified two mutations in the switch II region (G234A and E236A; Fig. 1) which provided further insight into the kinesin mechanism. G234 in kinesin corresponds to conserved glycines in the active sites of G proteins and myosin, which form a hydrogen bond with the γ -phosphate and trigger a conformational change between the NTP and NDP states⁴. Kinesin E236 is positioned farther away and does not directly contact the nucleotide. Kinetic analysis of these mutants and conformational studies (see below) provide evidence that neck linker docking to the catalytic core is necessary for kinesin stepping.

The G234A and E236A mutants did not move microtubules in a

gliding assay and had very low ATPase activities (<0.04 s⁻¹ in the presence of microtubules). The second order rate constants for the binding of mant ATP (a fluorescent ATP analogue) to mutant and wild-type kinesins were the same ($\sim 2 \times 10^6$ M⁻¹ s⁻¹). Unlike wild-type kinesin, however, both mutants did not display a phosphate burst after [γ -³²P]ATP was added to nucleotide-free proteins (data not shown), indicating that the hydrolysis step is rate-limiting and more than 1,000-fold slower than in wild-type kinesin.

Although G234A and E236A were similarly defective in ATP hydrolysis, further kinetic studies suggested that E236A, but not G234A, can take a single ATP-induced step on the microtubule. A two-step process has been described¹⁰⁻¹² that leads to ADP release from a kinesin dimer, which reveals communication between the two heads. After rapid mixing with microtubules, one head releases its ADP rapidly (100 s⁻¹; Table 2, reaction 2), whereas the second does so much more slowly (~ 0.4 s⁻¹)^{11,12}, indicating that it is constrained from interacting with the microtubule. ATP binding to the microtubule-bound head causes a conformational change that enables the detached partner head to step forward to the next tubulin binding site and release its ADP (Table 2, reaction 3). Both the G234A and E236A mutants in solution released mant ADP much faster than wild type, but microtubules further increased

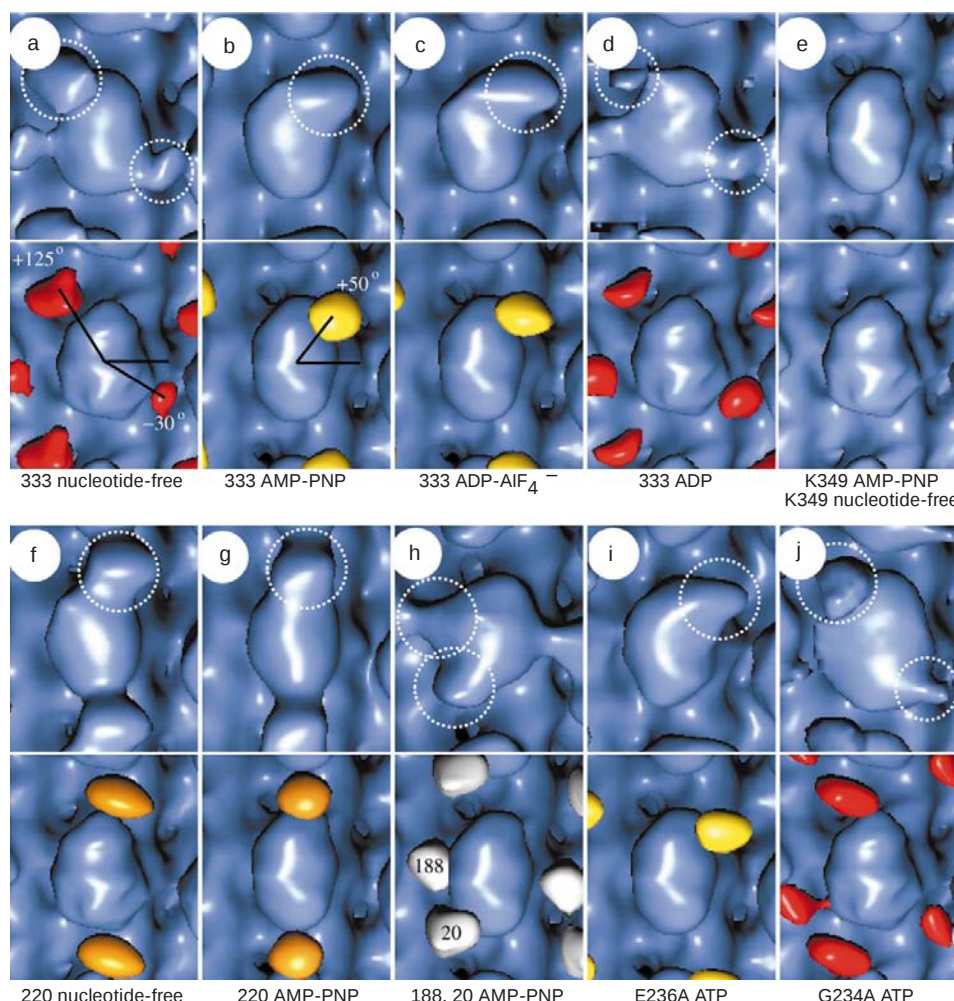


Figure 4 Three-dimensional and difference maps of gold labelled motors attached to microtubules calculated by cryo-EM and image analysis. Unlabelled controls used in the difference mapping are shown in **e** (see Methods). Density associated with a single motor domain is centred in each panel; additional densities in the corners of some panels are from adjacent motors on the microtubule surface lattice. The microtubule protofilament to which the motor is attached runs vertically with the plus end at the top. Density attributable to the gold particle is identified by dotted circles in the raw 3D maps and is

coloured in the difference maps. Difference maps are displayed with the control map that was used in the calculation. Red (**a, d, j**) and yellow (**b, c, i**) difference maps identify nucleotide states with ADP-like and ATP-like neck linker conformations respectively. Orange difference maps (**f, g**) show the position of the gold particle attached to residue 220 in two nucleotide states, which indicates that the catalytic core is immobile. The white difference map (**h**) shows gold particles attached to residues 20 and 188. The data in panels **g** and **h** confirm the docked orientation of the atomic model published earlier³¹.

these rates (Table 2). For E236A, ATP also accelerated ADP release from the second head by more than 100-fold, to the rate observed for wild-type kinesin (Table 2). As E236A cannot hydrolyse ATP on this timescale, this result confirms speculations based upon nucleotide analogues^{11,12} that ATP binding, and not a subsequent event in the cycle, causes the conformational change that leads to forward stepping. In contrast, ATP binding to the microtubule-bound G234A head caused only a 1.5-fold acceleration of ADP release by the second head, indicating a defect in head-head communication. Thus, our results show that E236A, but not G234A, can execute an ATP-binding-induced conformational change that allows the partner head to step to the next tubulin-binding site and release its ADP. These results also show interesting parallels with those obtained from two comparable ATP nonhydrolysing switch II mutations in *Dictyostelium* myosin (G457A and E459A), which are trapped in pre- and post-ATP 'power stroke' states, respectively¹⁶.

Neck linker docking in E236A but not G234A

If the ATP-driven conformational change in the neck linker is important for kinesin stepping, then ATP binding should be able to dock the neck linker onto the catalytic core in E236A, but not G234A. To test this hypothesis, we examined FRET between probes positioned at 220 and 333 in these two mutants. G234A showed no significant nucleotide- or microtubule-dependent changes in energy transfer efficiency; the values were comparable to that of nucleotide-free wild-type kinesin bound to microtubules (Table 1). However, a large FRET increase (a distance change of ~ 10 Å) was observed when microtubule-docked E236A bound either ATP or AMP-PNP. The comparable FRET signals obtained with ATP and AMP-PNP indicate that AMP-PNP mimics a normal nucleotide triphosphate state.

The neck linker conformations in the ATP state of the microtubule-bound G234A and E236A mutants were also examined by cryo-EM (Fig. 4). Consistent with the FRET results, the C333-attached gold particle on E236A-ATP was located at the $+50^\circ$ position at the tip of the catalytic core, similar to wild-type kinesin with AMP-PNP (Fig. 4h). For G234A-ATP, however, weak gold particle densities were detected at the $+125^\circ$ and -30° positions, but not at $+50^\circ$, which is similar to that observed for wild-type kinesin in the nucleotide-free and ADP states (Fig. 4g).

In conclusion, for E236A bound to microtubules, the neck linker is displaced from the tip of the catalytic core in the absence of nucleotide and redocks after ATP binding. Thus, E236A undergoes both ATP-induced neck linker docking and motor stepping to elicit ADP release (Tables 1, 2). In contrast, even with ATP in the active site, the neck linker of G234A is in an ADP-like state, and it cannot advance its partner head effectively to the next subunit. The G234A

result also reveals that hydrogen bonding of switch II to the γ -phosphate is necessary for triggering ATP-induced neck linker docking.

The kinesin force-generating cycle

Cryo-EM, EPR and FRET studies have provided complementary information for understanding how the neck linker changes its structure during kinesin's enzymatic cycle. Beginning a cycle with kinesin-ADP in solution, EPR and FRET suggest that the neck linker is not rigidly attached to the catalytic core. After kinesin binds to the microtubule and releases its ADP the neck linker is primarily detached, but, upon ATP binding, the neck linker rigidly docks onto the catalytic core in a single position ($+50^\circ$) that agrees with that seen in the crystal structure. This conformational change requires both microtubules and a γ -phosphate in the active site, as it is only observed with ATP, AMP-PNP and ADP- AlF_4^- . The last result also suggests that the docked state may be maintained in the ADP- P_i state in the normal enzymatic cycle. After phosphate release, the neck linker reverts to its more mobile conformation, and the

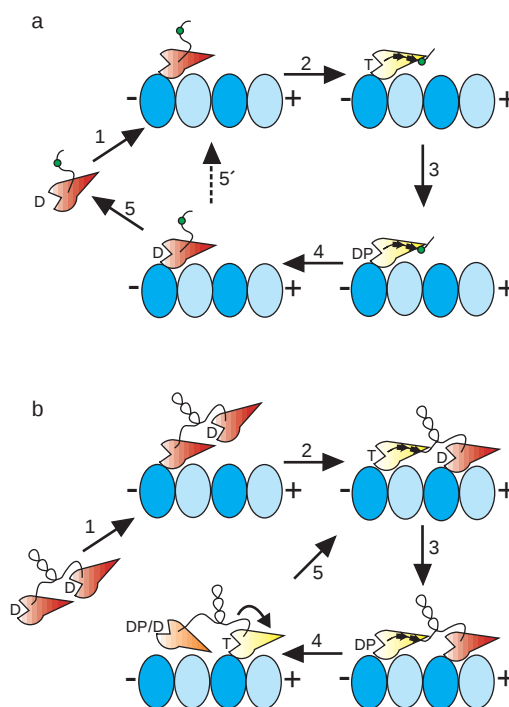


Figure 5 Models for motility by truncated kinesin monomers and processive dimers. Microtubule-bound, nucleotide triphosphate states are shown in yellow, and other states in red. The α - and β -subunits of tubulin are shown in two shades of blue. **a**, Kinesin monomers. Once bound to the microtubule, ATP (T) binding causes the neck linker to dock onto the catalytic core (step 2), which causes movement of position 333 (green circle) towards the microtubule plus end. Previous work has shown that, in solution, the monomer-ADP (D) state may not dissociate during each ATPase cycle^{37,38} (step 5'). This would reduce the velocity of motion; however, strain created by several active kinesin motors attached to the same microtubule may promote monomer-ADP detachment, leading to more efficient movement^{19,37,38} (step 5). **b**, Processive dimers. In the dimer cycle, one head initially attaches to the microtubule, releasing its ADP (step 1). ATP binding to this head causes neck linker docking and displacement of the partner head to a forward binding site (step 2). ATP is exchanged for ADP in the forward head (steps 3 and 4) causing the re-zipping of the neck linker. Detachment of the rear head (in either an ADP- P_i (DP) or ADP state) may be aided either by this re-zipping (as shown) or earlier by the weak to strong microtubule binding transition in the forward head. Once detached, the rear head would reattach to a forward binding site 16 nm away because of completed neck linker docking by the partner head (step 5). For kinesin walking down the microtubule, the cycle would continue through the three states shown (3, 4 and 5), and both heads would be microtubule bound for most of the time.

Table 2 Release of mant ADP from wild-type and mutant kinesin dimers (K420)

Reaction	1	2	3
Protein			
Wild type	0.01 ± 0.002 (1)	100 ± 10 (10,000)	120 ± 20 (12,000)
G234A	10 ± 1 (1)	40 ± 10 (4)	15 ± 3 (1.5)
E236A	1.1 ± 0.3 (1)	100 ± 15 (91)	120 ± 17 (109)

To determine the rate constant for reaction 1, K420-mant ADP (3 μM site concentration) was reacted with 500 μM ATP. For reaction 2, K420-mant ADP was reacted with a range of microtubule concentrations; the maximal rates are reported here. The amplitude of the fluorescence signal was roughly half that observed in the presence of ATP. For reaction 3, the K420-mant ADP was first mixed with 10–20 μM microtubules, and the one-head bound intermediate was then reacted with 500 μM ATP to promote the release of the remaining mant ADP.

* Fold rate increase compared with reaction 1 is given in parentheses.

motor binds more weakly to the microtubule³⁴. The above model is likely to apply to dimeric kinesin, although the neck coiled-coil and head-head contacts²⁶ may affect the conformation of the detached neck linker.

A model for how ATP-induced neck linker docking can drive kinesin motility is shown in Fig. 5. Kinesin binds to the microtubule with its catalytic core tip pointing towards the microtubule plus end (Figs 1 and 4). For the kinesin monomer (Fig. 5a), the docking of the neck linker from its various nucleotide-free positions to its single ATP position on the catalytic core would cause a plus-end-directed displacement. EPR results suggest that this transition might occur by a sequential zippering of the neck linker that begins at $\alpha 6$ and ends at the tip of the catalytic core. In this model, force generation does not occur by a 'power stroke' between two well-ordered states, as is generally described for myosin. Instead, movement involves a transition from a disordered to an ordered state, with ATP binding providing the energy source for rectifying this Brownian ratchet. This model also predicts that force generation in the monomer occurs upon ATP binding³⁵ and that multiple monomers, each producing small displacements, must interact simultaneously with a microtubule to produce movement^{7,9} (Fig. 5).

Neck linker movements could also explain processive motion by the kinesin dimer (Fig. 5b). Neck linker docking induced by ATP in the microtubule-bound head would position the partner head closer to the microtubule plus end²². Even if the neck coiled-coil remains intact, the partner head can then attach to the forward tubulin subunit³⁶. Once the two-head-bound intermediate is formed, ADP/ATP exchange in the forward head again triggers zippering of the neck linker, and this process could assist in dissociating the rear head^{9,37,38} and displacing it by 16 nm to the next available forward binding site. The net result of this action would be a rapid 8-nm step⁶. The bias for plus-end-directed motion in the kinesin dimer is generated through both neck linker docking and microtubule binding by the forward head. In this model, neck linker movement provides the trigger that allows motor stepping, although changes in the neck coiled-coil or in head-head interactions may augment this process^{33,36}. The model also suggests that a backward load could slow down the motor³⁹ by decreasing the rate of neck linker docking and cause kinesin to take a backward step at stall loads⁴⁰ by reversing this process. A forward load could speed up movement⁴⁰ by favouring neck linker docking and/or by accelerating the dissociation of the rear head.

Comparison with other motors

The motility-producing conformational change described here for kinesin differs significantly from the rotation of myosin's rigid lever arm which is triggered by phosphate release^{13,14,30}. In both motors, however, the communication mechanism from the active site to the mechanical element probably operates through structurally and functionally analogous nucleotide-binding loops and helices³ (for example, switch II, $\alpha 4$ and $\alpha 6$; Fig. 1). In addition, the neck linker of kinesin emerges from the catalytic core in a similar position to the reactive-sulphydryl helix of myosin³, which is also thought to undergo a disordered-to-ordered transition during the ATPase cycle of myosin^{30,41}. Myosin's reactive sulphydryl helix is believed to be essential for driving a 'rearward-to-forward' motion of the converter-lever arm complex, similar to the rearward-to-forward movement of the partner kinesin head triggered by the neck linker. Thus, although differently engineered, the converter-lever-arm complex of myosin, and the second head-neck coiled-coil of kinesin may be analogous structures that respond to a similarly positioned 'trigger' peptide.

Our proposed kinesin mechanism also reveals similarities to other mechanically active enzymes. Kinesin's motility-producing conformational change is driven by ATP binding, but this nucleotide-dependent conformational change requires motor contact with its microtubule substrate. Similarly, in F_1 -ATP synthase^{42,43} and

GroEL⁴⁴, functionally relevant conformational changes are driven by ATP binding while these enzymes are complexed with their partner proteins (the γ subunit for F_1 and GroES for GroEL). A simultaneous requirement of ATP and the partner protein for triggering a conformational change ensures tight coupling between ATP-binding energy and mechanical work in all of these enzymes. □

Methods

Cloning and protein purification

Solvent-exposed cysteines were mutated to Ser/Ala (C7S, C65A, C168A, C174S, C294A, C330S and C421A) by QuikChange mutagenesis (Stratagene) of a human K560-His₆ construct in a pET17b vector¹⁷. Buried cysteines at positions 13, 301 and 302 were retained. This 'cys-light' kinesin was unreactive to thiol-reactive agents, exhibited normal microtubule gliding velocity (cys-light K560 and wild-type K560: 19.3 ± 2.8 and $20.2 \pm 3.6 \mu\text{m min}^{-1}$, respectively) and microtubule-stimulated ATPase activity (cys-light K560 and wild-type K560: catalytic rate constant (k_{cat}) $\sim 19.3 \pm 1.5$ and $28.6 \pm 4.2 \text{ ATP head}^{-1} \text{ s}^{-1}$, respectively; Michaelis constant (K_m) $\sim 1.2 \pm 0.5$ and $1.2 \pm 0.6 \mu\text{M}$ tubulin, respectively), and moved processively as determined using a single molecule fluorescence assay¹⁷. Other constructs described in this study were prepared using QuikChange mutagenesis and/or PCR subcloning and were confirmed by DNA sequencing.

Proteins were expressed in BL21(DE3)¹⁷, and K560¹⁷, K420²⁷ and kinesin-GFP fusion proteins¹⁷ were purified as previously described. Purification of cys-light K349 was done as described¹⁵, except that HEPES buffer was used instead of PIPES, and a final Resource Q column was run at pH 8 with 0.5 mM Tris-carboxyethyl phosphine-HCl (TCEP) instead of dithiothreitol. Proteins were frozen in 20% w/v sucrose and stored in liquid nitrogen.

Enzymology and motility assays

Steady-state, microtubule-stimulated ATPase rates were measured using a coupled enzymatic assay⁴⁶ or by hydrolysis of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ ³⁴. Dissociation kinetics of mant ADP from kinesin was done as described³⁴ in 25 mM PIPES, pH 6.9, 2 mM MgCl₂, 1 mM EGTA and 50 mM NaCl.

Tubulin preparation and microtubule gliding assays for K560 proteins were done as described⁴⁶. Monomer motility was measured in K339-GFP using anti-GFP antibodies to attach the protein to the surface¹⁷. The motility was assayed in FRET labelling buffer (see below) with 0.5 mM TCEP-HCl, 2 mg ml⁻¹ casein, 10 $\mu\text{g ml}^{-1}$ unlabelled microtubules, 20 μM taxol, an oxygen depleting system and 1 mM ATP; microtubules were visualized using DIC microscopy, and more than 25 velocity measurements were made. For testing the motility of chemically modified kinesins, probes were labelled to $\sim 100\%$ stoichiometry as confirmed by mass spectroscopy, fluorescence or gel-shift assays. The microtubule gliding velocities produced by MSL-C333 K339-GFP, TMR/CPM-C220/C333 K339-GFP, TMR-C220 K339-GFP and undecagold-C333 K339-GFP were 107.8%, 81.9%, 98.9% and 86.6% of the cys-light K339-GFP control, which moved at $2.4 \mu\text{m min}^{-1}$.

Chemical modification of kinesin

For EPR, a single cysteine-containing kinesin in EPR labelling buffer (25 mM PIPES, pH 7, 50 mM NaCl, 2 mM MgCl₂, 1 mM EGTA and 10 μM ATP) was incubated with a 2–5-fold molar excess of 4-maleimido-2,2,6,6-tetramethyl-1-piperidinoxy (MSL; Sigma) overnight at 4 °C. Excess spin label was removed as described²⁴. The labelling stoichiometry was determined by measuring protein concentration (Bradford assay with BSA standards) and probe concentration (by comparing labelled protein to known concentrations of MSL²⁴).

For FRET, double-cysteine containing kinesin (K349) was dialysed against FRET labelling buffer (EPR buffer at pH 7.5 with 0.1 mM TCEP) and reacted with 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM, Molecular Probes) for 3 h at room temperature at a stoichiometry of ~ 1 dye per 10 reactive cysteines. Half of the CPM-labelled kinesin (or K339-GFP) was removed (donor-labelled kinesin), and the remainder (donor-acceptor-labelled kinesin) was incubated for 3 h at room temperature with a 10-fold excess of tetramethylrhodamine-6-maleimide to cysteines (TMR, Molecular Probes) which reacted with the remaining reactive sites. Thus, almost all donor-labelled kinesins were modified with an acceptor probe. Removal of free dye was done as described²⁴. The labelling stoichiometry was determined by measuring dye absorbance and protein concentration⁴⁶.

For gold-cluster labelling, monomaleimide-undecagold labelling was carried out for 16 h at 4–5-fold excess⁴⁷. Unbound gold-cluster and aggregated material were removed by ion exchange and gel filtration chromatography. Undecagold labelling shifted the mobility of kinesin in non-denaturing PAGE; the extent of labelling assessed by this method was 80–100%.

Spectroscopy

Electron-paramagnetic resonance spectra were obtained, and the ΔH was determined as described²⁴. For EPR, kinesin was incubated with either 5 mM Mg-AMP-PNP, 5 mM Mg-ADP, 2 mM Mg-ADP-AlF₄ (2 mM AlCl₃, 10 mM KF), or apyrase (6.6 U ml⁻¹) (nucleotide-free). In solution, 50–80 μM kinesin was used. For kinesin-microtubule samples, 25–30 μM kinesin was mixed with 125–150 μM polymerized tubulin, the complex centrifuged, and microtubule pellet analysed to ensure that essentially all kinesin was bound to microtubules.

Fluorescence resonance energy transfer samples contained $\sim 1 \mu\text{M}$ kinesin (in FRET labelling buffer), $\pm 10 \mu\text{M}$ microtubules, 5 mM of the indicated magnesium nucleotide or 25 U ml^{-1} apyrase to generate the nucleotide-free state. The microtubule-bound, ADP state and the solution, nucleotide-free state could not be measured because of excessive light scattering from high microtubule concentrations and protein denaturation, respectively. Using a K2 spectrofluorimeter (ISS), fluorescence was measured at two serial twofold dilutions to ensure that energy transfer was not dilution sensitive because of inner filter effects or reabsorption of donor emission. GFP and CPM were excited at 475 nm and 395 nm respectively, and TMR emission was measured at 575 nm. Energy transfer to the nearby acceptor was measured by the reduction in donor fluorescence ($E = 1 - (F_{\text{DA}}/F_{\text{DO}})$), where F_{DO} and F_{DA} are the donor fluorescence in the absence and presence of the acceptor, respectively⁴⁸). Energy transfer for K339–GFP–TMR samples was calculated both by donor quenching and acceptor enhancement⁴⁸; both methods yielded similar values. Donor–acceptor distance was calculated from Förster's equation, assuming orientation parameter κ^2 to be two-thirds (ref. 49), a quantum yield of 0.5 for coumarin⁴⁸, and a spectral overlap for coumarin-tetramethylrhodamin of $6.4 \times 10^{14} \text{ M}^{-1} \text{ cm}^{-1} \text{ nm}^4$ (ref. 48).

Cryo-EM and image analysis

Frozen grids of decorated microtubules were prepared and imaged as described³¹ with some modifications. Microtubule decoration with gold-labelled kinesin heads ($2\text{--}10 \text{ mg ml}^{-1}$) was carried out in 80 mM PIPES (pH 6.8), 1 mM MgCl_2 , 1 mM EGTA supplemented with 5 mM magnesium nucleotide or apyrase ($100\text{--}200 \text{ U ml}^{-1}$) (nucleotide-free). Images of helical 15-protofilament microtubules were analysed as described³¹. Two unlabelled control maps (K349 with AMP-PNP and nucleotide-free K349) were used for difference mapping. Subtle structural features of the core regions of the raw gold labelled 3D maps dictated which control map was used for the difference calculation. Thus, difference maps of nucleotide-free and ADP-containing complexes and of G234A with ATP were calculated using the nucleotide-free control map. The remaining difference maps were calculated using the K349-AMP-PNP control map. The number of data sets averaged and the total number of molecules for each of the raw 3D maps shown in Fig. 4 were 22 data sets, 10,780 molecules (a); 42, 21, 280 (b); 18, 7420 (c); 25, 16,380 (d); 28, 18,620 (e, top); 28, 12,460 (e, bottom); 14, 6,160 (f); 18, 10,360 (g); 16, 7,980 (h); 11, 10,500 (i); and 18, 6,440 (j).

Received 9 September; accepted 11 November 1999.

- Goldstein, L. S. B. & Philp, A. V. The road less traveled: emerging principles of kinesin motor utilization. *Annu. Rev. Cell Dev. Biol.* **15**, 141–183 (1999).
- Vale, R. D. & Fletterick, R. J. The design plan of kinesin motors. *Annu. Rev. Cell Dev. Biol.* **13**, 745–777 (1997).
- Kull, F. J., Sablin, E. P., Lau, R., Fletterick, R. J. & Vale, R. D. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* **380**, 550–555 (1996).
- Sablin, E. P., Kull, F. J., Cooke, R., Vale, R. D. & Fletterick, R. J. Crystal structure of the motor domain of the kinesin-related motor ncd. *Nature* **380**, 555–559 (1996).
- Howard, J., Hudspeth, A. J. & Vale, R. D. Movement of microtubules by single kinesin molecules. *Nature* **342**, 154–158 (1989).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Berliner, E., Young, E. C., Anderson, K., Mahtani, H. & Gelles, J. Failure of a single-headed kinesin to track parallel to microtubule protofilaments. *Nature* **373**, 718–721 (1995).
- Vale, R. D. *et al.* Direct observation of single kinesin molecules moving along microtubules. *Nature* **380**, 451–453 (1996).
- Hancock, W. O. & Howard, J. Processivity of the motor protein kinesin requires two heads. *J. Cell Biol.* **140**, 1395–1405 (1998).
- Hackney, D. D. Evidence for alternating head catalysis by kinesin during microtubule-stimulated ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **91**, 6865–6869 (1994).
- Ma, Y.-Z. & Taylor, E. W. Interacting head mechanism of microtubule-kinesin ATPase. *J. Biol. Chem.* **272**, 724–730 (1997).
- Gilbert, S. P., Moyer, M. L. & Johnson, K. A. Alternating site mechanism of the kinesin ATPase. *Biochemistry* **37**, 792–799 (1998).
- Rayment, I. *et al.* Structure of the actin–myosin complex and its implications for muscle contraction. *Science* **261**, 58–65 (1993).
- Dominguez, R., Freyzo, Y., Trybus, K. M. & Cohen, C. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-power stroke state. *Cell* **94**, 559–571 (1998).
- Corrie, J. E. T. *et al.* Dynamic measurements of myosin light-chain-domain tilt and twist in muscle contraction. *Nature* **400**, 425–430 (1999).
- Suzuki, Y., Yasunaga, T., Ohkura, R., Wakabayashi, T. & Sutoh, K. Swing of the lever arm of a myosin motor at the isomerization and phosphate-release steps. *Nature* **396**, 380–383 (1998).
- Case, R. B., Pierce, D. W., Hom-Booher, N., Hart, C. L. & Vale, R. D. The directional preference of kinesin motors is specified by an element outside of the motor catalytic domain. *Cell* **90**, 959–966 (1997).
- Henningsen, U. & Schliwa, M. Reversal in the direction of movement of a molecular motor. *Nature* **389**, 93–96 (1997).
- Endow, S. A. & Waligora, K. W. Determinants of kinesin motor polarity. *Science* **281**, 1200–1202 (1998).
- Arnal, I. & Wade, R. H. Nucleotide-dependent conformation of the kinesin dimer interacting with microtubules. *Structure* **1998**, 33–38 (1998).
- Hirose, K., Lowe, J., Alonso, M., Cross, R. A. & Amos, L. A. Congruent docking of dimeric kinesin and ncd into three-dimensional electron cryomicroscopy maps of microtubule-motor ADP complexes. *Mol. Biol. Cell* **10**, 2063–2074 (1999).
- Hirose, K., Lockhart, A., Cross, R. A. & Amos, L. A. Three-dimensional cryoelectron microscopy of dimeric kinesin and ncd motor domains on microtubules. *Proc. Natl Acad. Sci. USA* **93**, 9539–9544 (1996).
- Mchaurah, H. S., Lietzow, M. A., Hideg, K. & Hubbell, W. L. Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry* **35**, 7692–7704 (1996).
- Naber, N., Cooke, R. & Pate, E. Binding of ncd to microtubules induces a conformational change near the junction of the motor domain with the neck. *Biochemistry* **36**, 9681–9689 (1997).
- Sack, S. *et al.* X-ray structure of motor and neck domains from rat brain kinesin. *Biochemistry* **36**, 16155–16165 (1997).
- Kozielewski, F. *et al.* The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. *Cell* **91**, 985–994 (1997).
- Stone, D. B., Hjelm, R. P. Jr & Mendelson, R. A. Solution structures of dimeric kinesin and Ncd motors. *Biochemistry* **38**, 4938–4947 (1999).
- Marx, A. *et al.* Conformations of kinesin: solution versus crystal structures and interactions with microtubules. *Eur. Biophys. J.* **27**, 455–465 (1998).
- Fisher, A. J. *et al.* X-ray structures of the myosin motor domain of *Dictyostelium discoideum* complexed with MgADP.BcFx and MgADP.AIF₃⁻. *Biochemistry* **34**, 8960–8972 (1995).
- Houdusse, A., Kalabokis, V. N., Himmel, D., Szent-Gyorgyi, A. G. & Cohen, C. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head. *Cell* **97**, 459–470 (1999).
- Sosa, H. *et al.* A model for the microtubule-ncd motor protein complex obtained by cryo-electron microscopy and image analysis. *Cell* **90**, 217–224 (1997).
- Kozielewski, F., Arnal, I. & Wade, R. H. A model of the microtubule-kinesin complex based upon electron cryomicroscopy and X-ray crystallography. *Curr. Biol.* **8**, 191–198 (1998).
- Hoenger, A. *et al.* Image reconstructions of microtubules decorated with monomeric and dimeric kinesins: comparison with X-ray crystallography. *Curr. Biol.* **8**, 191–198 (1998).
- Ma, Y. Z. & Taylor, E. W. Mechanism of microtubule kinesin ATPase. *Biochemistry* **34**, 13242–13251 (1995).
- Romberg, L. & Vale, R. D. Chemomechanical cycle of kinesin differs from that of myosin. *Nature* **361**, 168–170 (1993).
- Romberg, L., Pierce, D. W. & Vale, R. D. Role of the kinesin neck region in processive microtubule-based motility. *J. Cell Biol.* **140**, 1407–1416 (1998).
- Jiang, W. & Hackney, D. D. Monomeric kinesin head domains hydrolyze multiple ATP molecules before release from a microtubule. *J. Biol. Chem.* **272**, 5616–5621 (1997).
- Moyer, M. L., Gilbert, S. P. & Johnson, K. A. Pathway of ATP hydrolysis by monomeric and dimeric kinesin. *Biochemistry* **37**, 800–813 (1998).
- Visscher, K., Schnitzer, M. J. & Block, S. M. Single kinesin molecules studied with a molecular force clamp. *Nature* **400**, 184–189 (1999).
- Coppin, C. M., Pierce, D. W., Hsu, L. & Vale, R. D. The load dependence of kinesin's mechanical cycle. *Proc. Natl Acad. Sci. USA* **94**, 8539–8544 (1997).
- Thomas, D. D., Ramachandran, S., Roopnarine, O., Hayden, D. W. & Ostap, E. M. The mechanism of force generation in myosin: a disordered-to-order transition, coupled to internal structural changes. *Biophys. J.* **68** (suppl.), 135s–141s (1995).
- Wang, H. & Oster, G. Energy transduction in the F1 motor of the ATP synthase. *Nature* **396**, 279–282 (1998).
- Al-Shawi, M. & Nakamoto, R. Mechanism of energy coupling in the FoF1-ATP synthase: the uncoupling mutant, γM23K , disrupts the use of binding energy to drive catalysis. *Biochemistry* **36**, 12954–12960 (1997).
- Weissman, J. S., Rye, H. S., Fenton, W. A., Beechem, J. M. & Horwich, A. L. Characterization of the active intermediate of a GroEL–GroES-mediated protein folding reaction. *Cell* **84**, 481–490 (1996).
- Fujiwara, S., Kull, F. J., Sablin, E. P., Stone, D. B. & Mendelson, R. A. The shapes of the motor domains of two oppositely directed microtubule motors, ncd and kinesin: a neutron scattering study. *Biophys. J.* **69**, 1563–1568 (1995).
- Woehlke, G. *et al.* Microtubule interaction site of the kinesin motor. *Cell* **90**, 207–216 (1997).
- Safer, D. Undecagold cluster labeling of proteins at reactive cysteine residues. *J. Struct. Biol.* **127**, 371–374 (1999).
- Van Der Meer, B. W., Coker, G. & Chen, S.-Y. S. *Resonance Energy Transfer: Theory and Data* (VCH, New York, 1994).
- dos Remedios, C. G. & Moens, P. D. Fluorescence resonance energy transfer spectroscopy is a reliable “ruler” for measuring structural changes in proteins. Dispelling the problem of the unknown orientation factor. *J. Struct. Biol.* **115**, 175–185 (1995).
- Barnett, V. A., Fajer, P., Polnaszek, C. F. & Thomas, D. D. High resolution detection of muscle crossbridge orientation by EPR. *Biophys. J.* **49**, 144–147 (1986).

Acknowledgements

We thank L. Sweeney for encouragement; A. Ruby for assistance with protein preparations; B. Sheehan for computing and image processing; and M. Tomishige for examining the processivity of cys-light K560–GFP. We thank R. Case and K. Thorn for comments on the manuscript. This work was supported, in part, by grants from the NIH (R.A.M., E.W.T., R.C. and R.D.V.), and the NSF (B.O.C.). S.R. is supported by the UCSF Graduate Group in Biophysics.

Correspondence and requests for materials should be addressed to R.D.V. (e-mail: vale@phy.usc.edu).

Thermally activated transitions in a bistable three-dimensional optical trap

Lowell I. McCann†, Mark Dykman & Brage Golding

Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824-1116, USA

Activated escape from a metastable state underlies many physical, chemical and biological processes: examples include diffusion in solids, switching in superconducting junctions^{1,2}, chemical reactions^{3,4} and protein folding^{5,6}. Kramers presented the first quantitative calculation⁷ of thermally driven transition rates in 1940. Despite widespread acceptance of Kramers' theory⁸, there have been few opportunities to test it quantitatively as a comprehensive knowledge of the system dynamics is required. A trapped brownian particle (relevant to our understanding of the kinetics, transport and mechanics of biological matter^{9,10}) represents an ideal test system. Here we report a detailed experimental analysis of the brownian dynamics of a sub-micrometre sized dielectric particle confined in a double-well optical trap. We show how these dynamics can be used to directly measure the full three-dimensional confining potential—a technique that can also be applied to other optically trapped objects^{11,12}. Excellent agreement is obtained between the predictions of Kramers' theory and the measured transition rates, with no adjustable or free parameters over a substantial range of barrier heights.

A mesoscopic particle, suspended in a liquid and confined within a metastable potential well provides an ideal representation of Kramers' ideas. The particle moves at random within the well until a large fluctuation propels it out of the well over an energy barrier. The potential can be created with a gradient optical trap—a technique widely used in biophysical studies⁹. Dual optical traps were introduced initially to study the synchronization of the interwell transitions by periodic forcing¹³. A particle in a dual optical trap is a well-controlled model system which can be used to address quantitatively the problem of transition rates provided the confining potential can be accurately determined.

The optically induced potential wells constructed in the present experiments are formed by focusing two parallel laser beams through a single objective lens. Each beam creates a stable three-dimensional trap as a result of electric field gradient forces exerted on a transparent dielectric spherical silica particle of diameter $2R = 0.6 \mu\text{m}$ (Bangs Laboratories). Displaced by 0.25 to $0.45 \mu\text{m}$, the beams create a double-well potential, with the stable positions of the particle centre at $\mathbf{r}_1$ and $\mathbf{r}_2$. Relatively infrequent random transitions between the potential wells occur through a saddle point at $\mathbf{r}_s$ as depicted in Fig. 1a. Both the depth of each potential well and the height of the intervening barrier are under experimental control.

The two HeNe lasers (17 mW, 633 nm) that create the traps are stabilized by Pockels cell electro-optic modulators and imaged into a sample cell by a 100x 1.4 NA PlanApo objective lens mounted in an Olympus IX-70 microscope. The beams are mutually incoherent and circularly polarized as they enter the microscope. A single trapped sphere is imaged onto a Dalsa CA-D1 CCD camera operated at 200 frames per second. The coordinates of the sphere's centre in the focal plane of the objective lens are found to within $\pm 10 \text{ nm}$ with a pattern-matching routine, and the coordinate in the beams' propagation direction is extracted by analysis of the image size as it moves above and below the focal plane. The sealed sample

cell, constructed from two glass coverslips and epoxy resin, holds a dilute suspension of silica spheres in water at room temperature. The experimental output of this system is a time record of the coordinates $\mathbf{r}(t)$; as shown in Fig. 1b.

As a result of the short equilibration time of the sphere in water ($\gamma^{-1} = M/(6\pi\eta R) \sim 10^{-7} \text{ s}$, where η is the viscosity of water and M is the particle mass), the brownian particle relaxes to equilibrium on a timescale much shorter than the sampling time. The spatial probability density is therefore:

$$\rho(\mathbf{r}) = Z^{-1} \exp(-U(\mathbf{r})/k_B T) \quad (1)$$

Here $U(\mathbf{r})$ is the potential energy as a function of particle position, and Z is a normalization constant. The probability density is found from the time series $\mathbf{r}(t)$, typically using 10^6 to 10^7 frames, for durations much longer than the mean interwell transition times. Equation (1) then allows us to obtain the potential directly from measurements.

Figure 2 shows $U(\mathbf{r})$ for a two-beam trap. We choose the x axis to be in the direction from one beam to the other and the z axis along the propagation direction of the beams. The potential minima, $\mathbf{r}_1$ and $\mathbf{r}_2$, lie in the symmetry plane $y = 0$ formed by the beam axes. Figure 2a shows a two-dimensional cross-section, at $y=0$, of the potential with energy contours at $1.0 k_B T$ intervals distinguished by colour-coding, where k_B is the Boltzmann constant and T is temperature. If, for a given x , we find the minimum of $U(\mathbf{r})$ over y and z , we obtain the familiar one-dimensional representation of a double-well potential shown in Fig. 2b. Figure 2c shows the energy contours at $\mathbf{r}_2$ for a cross-section in the $y-z$ plane with the corresponding energy profile in Fig. 2d. The elongated profile along z is expected, since the radial field gradient is determined by the transverse beam profile whereas the weaker axial gradient depends on the angular divergence of the focused beam near its diffraction-limited waist¹⁴.

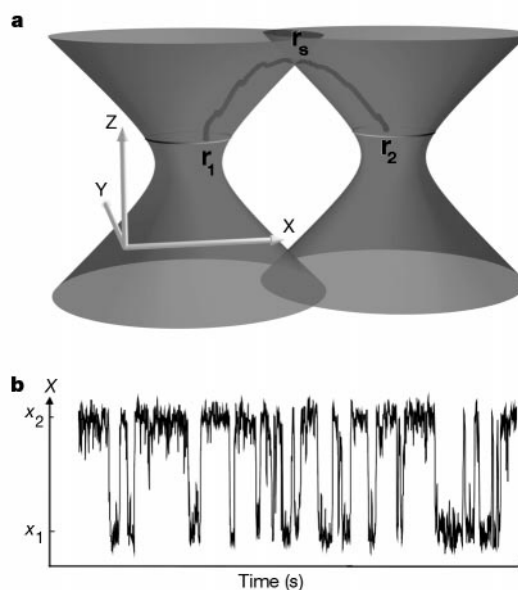


Figure 1 Interwell transitions in a dual optical trap. **a**, Two neighbouring focused beams in the absence of a perturbing trapped sphere are shown. The dark line illustrates the path of a trapped sphere in an interwell transition between $\mathbf{r}_1$ and $\mathbf{r}_2$ through the saddle at $\mathbf{r}_s$. The equilibrium positions $\mathbf{r}_1$ and $\mathbf{r}_2$, at the level of the rings around the beams, are displaced above the focal plane of the incident beams. **b**, Projection of particle trajectory on the x -axis perpendicular to the beams, where x_1, x_2 are components of $\mathbf{r}_1, \mathbf{r}_2$. The sampling interval is 5 ms and the total time duration of the record is approximately 8 s. During the acquisition time, a computer performs a pattern-matching routine that returns the three particle coordinates. The particle spends most of its time in the vicinity of the stable points $\mathbf{r}_1$ and $\mathbf{r}_2$ with infrequent transitions between them.

† Present address: Department of Physics, University of Wisconsin-River Falls, River Falls, Wisconsin 54022, USA.

Table 1 Features of the optical potential in Fig. 2 and associated transition rates

	$ \omega^{(1)} $	$\omega^{(2)}$	$\omega^{(3)}$	$\Delta U/k_B T$	W_0^K	W^K	W^{meas}
Well at $\mathbf{r}_1$	16 ± 4	55 ± 14	6 ± 3	2.77 ± 0.05	80 ± 40	5 ± 2	6.57 ± 0.03
Saddle at $\mathbf{r}_s$	7 ± 2	50 ± 25	6 ± 3				
Well at $\mathbf{r}_2$	21 ± 5	56 ± 14	6 ± 3	3.51 ± 0.04	110 ± 50	3 ± 2	3.70 ± 0.02

Characteristic frequencies $\omega_j^{(i)}$ (in 10^4 s^{-1} , with $j = 1, 2, 3$) associated with the extrema, $\mathbf{r}_i$, of the potential in Fig. 2. Note that $\omega_s^{(1/2)} < 0$. W_0^K is the Kramers prefactor, W^K is the Kramers rate, and W^{meas} is the experimentally determined rate in transitions per second, reported for the two potential wells at $\mathbf{r}_1$ and $\mathbf{r}_2$. $\Delta U/k_B T$ is the reduced barrier height. The errors reported for $\omega^{(3)}$ do not include the uncertainty in the z -position calibration since it does not influence the value of W_0^K .

The double-well potential in Fig. 2 has minima $\mathbf{r}_{1,2}$ separated by $\delta x = 0.35 \mu\text{m}$, a single intervening saddle point at $\mathbf{r}_s$, and $U(\mathbf{r}_2) - U(\mathbf{r}_1) = 0.8 k_B T$. We emphasize that the potential measured in this manner is the overall effective potential experienced by the particle in its environment.

A feature of the effective potential evident in Fig. 2a is the strong symmetry breaking about the focal plane, which is the symmetry plane of the beams, unperturbed by the particle. This symmetry breaking leads to the single saddle point in $U(\mathbf{r})$ instead of two saddle points as might be inferred from Fig. 1a. This aspect of the potential is not an artefact of specific experimental conditions, such as non-parallel optical beams, but is a consequence of the beam-particle interaction. The dielectric particle acts as a spherical lens to refocus the beam inside the sphere. When the particle is displaced in the $+z$ direction above the focal plane, the electromagnetic field is most strongly ‘squeezed’ into the particle, thus minimizing the total free energy of the polarized particle in the field.

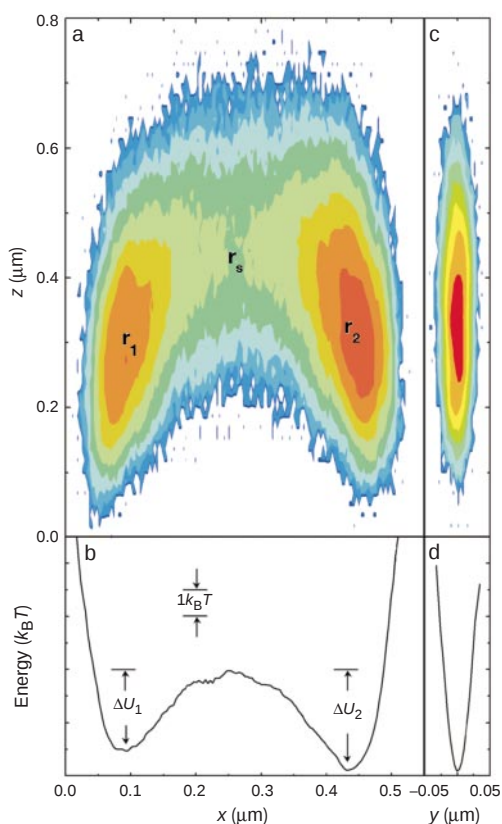


Figure 2 Experimentally determined potential energy of the particle in a double-well optical trap. **a**, Energy contours for a cross-section in the x - z plane that contains the stable points $\mathbf{r}_1$ and $\mathbf{r}_2$, and the saddle point $\mathbf{r}_s$. Each colour indicates a $1.0 k_B T$ energy interval. **b**, The energy, minimized with respect to y and z , as a function of x as shown in **a**. ΔU_1 and ΔU_2 are the barrier heights for the corresponding wells. **c**, The energy contours in a y - z cross-section containing point $\mathbf{r}_2$. **d**, The energy, minimized with respect to x and z , as a function of y as shown in **c**. The potential described in the figure was extracted from a data set of 4×10^6 camera frames containing 94,000 interwell transitions.

In the vicinity of $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_s$, the potential $U(\mathbf{r})$ is quadratic in the displacements $\delta \mathbf{r} = \mathbf{r} - \mathbf{r}_i$ with $i=1,2$, or s :

$$U(\mathbf{r}) = U(\mathbf{r}_i) + \frac{1}{2} M \sum_{\alpha, \beta} \Lambda_i^{\alpha\beta} \delta r_\alpha \delta r_\beta \quad (2)$$

Here $\alpha, \beta = x, y, z$. In practice, we perform a least-squares fit of equation (2) to the data in the vicinity of $\mathbf{r}_i$. As an example of the results and errors in this procedure, Table 1 shows the eigenvalues ω_i^2 of the matrix Λ_i for the potential shown in Fig. 2. The characteristic frequencies $|\omega_i|$ are small compared to the damping rate γ so the particle is overdamped.

We now consider the interwell dynamics of the particle. Specifically, we examine the rates W_{12} (W_{21}) of transitions $1 \rightarrow 2$ ($2 \rightarrow 1$), and their dependence on the energy barriers $\Delta U_1 = U(\mathbf{r}_s) - U(\mathbf{r}_1)$ and $\Delta U_2 = U(\mathbf{r}_s) - U(\mathbf{r}_2)$. We tune $\Delta U_i/k_B T$ up to 8.5 by adjusting the optical intensity in each beam and the beam separation. The presence of the energy asymmetry $\Delta U_{12} = U(\mathbf{r}_2) - U(\mathbf{r}_1)$ between 0.5 – $3 k_B T$ allows us to measure two independent rates with a single optical configuration. For each set of experimental conditions, we determine the full three-dimensional potential, similar to Fig. 2. To obtain satisfactory statistics, we accumulated between 2,400 and 94,000 interwell transitions that occurred over 5×10^5 to 10^7 frames.

A quantitative description of thermally activated escape from a one-dimensional metastable potential was given by Kramers⁷ and subsequently extended to multidimensional potentials¹⁵. The subject has been extensively reviewed^{8,16,17}. For an overdamped brownian particle in a potential $U(\mathbf{r})$, the Kramers transition rate is:

$$W^K = W_0^K \exp\left(-\frac{\Delta U}{k_B T}\right) \quad (3)$$

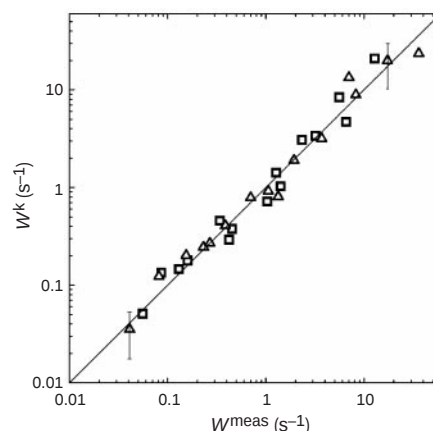


Figure 3 Experimental and theoretical transition rates. We compare the directly measured transition rates, W^{meas} , with the rates calculated from the three-dimensional Kramers theory, W^K , using the measured curvatures of the potential wells. The squares represent escapes from the well at $\mathbf{r}_1$ and the triangles represent escapes from the well at $\mathbf{r}_2$. The line of slope unity indicates the result expected if the data coincide with the Kramers theory.

Here the prefactor is given by the following expression for a three-dimensional potential¹⁵:

$$W_0^K = \frac{|\omega_s^{(1)}| \omega_s^{(1)} \omega_s^{(2)} \omega_s^{(3)}}{2\pi\gamma} \quad (4)$$

Here $\omega_s^{(j)}$ and $\omega_s^{(j)}$ characterize, respectively, the curvatures of the potential at the saddle point and at the minimum from which the system escapes, with $(\omega_s^{(1)})^2 < 0$. Therefore, with knowledge of the potential, not only the exponential term, but also the prefactor can be explicitly computed, as shown in Table 1.

Figure 3 shows a plot of the Kramers rates, W^K , calculated from equations (3) and (4) as a function of the transition rates, W^{meas} , obtained from the mean dwell time in each state or by fitting an exponential to a histogram of dwell times, in accordance with a Poisson distribution, which yielded equivalent results. The systematic variation of the potential barrier ΔU by approximately $6 k_B T$ is responsible for the nearly three-decade variation in transition rates. The solid line with unity slope denotes the coincidence of theory and experiment. The data fall remarkably close to the line, confirming the multidimensional Kramers theory of transition rates.

One of the major contributors to the uncertainty in calculating W^K is the error in the saddle point frequencies entering equation (4). This is primarily statistical, since the particle spends little time in the vicinity of r_s . This error is amplified for the highest barriers with the lowest transition rates. At low barriers, the statistical uncertainty is small since the transition rates are large, but the Kramers theory is not strictly valid here. If the thermal diffusion length exceeds the size of the parabolic region in the vicinities of the stationary states, then the transition rates are affected by the shape of the potential away from the stationary states. However, the corrections to the theory are small in the range of parameters studied here.

Detailed knowledge of the overall potential, as afforded by our experiments, should enable investigations of the escape rate of underdamped particles in the region in which escape occurs by diffusion over energy, as well as the Kramers turnover region^{8,18}. This can be accomplished with present methods by reducing the viscous damping on the particle. Knowledge of the potential is also crucial for understanding strategies for control of escape. Escape occurs by large fluctuations¹⁹ that move the system from a minimum of the potential to the barrier top along optimal paths. The results of this work make it possible to find these paths, enabling selective control of escape rates by external modulating fields²⁰.

Received 12 August; accepted 7 October 1999.

- Han, S., Lapoint, J. & Lukens, J. E. Effect of a two-dimensional potential on the rate of thermally induced escape over the potential barrier. *Phys. Rev. B* **46**, 6338–6345 (1992).
- Devoret, M. H., Esteve, D., Martinis, J. M., Cleland, A. & Clarke, J. Resonant activation of a brownian particle out of a well: microwave-enhanced escape from the zero-voltage state of a Josephson junction. *Phys. Rev. B* **36**, 58–73 (1987).
- van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1992).
- Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Chem. Phys.* **81**, 2340–2361 (1977).
- Sali, A., Shakhnovich, E. & Karplus, M. How does a protein fold? *Nature* **369**, 248–251 (1994).
- White, S. H. & Wimley, W. C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 319–365 (1999).
- Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).
- Melnikov, V. I. The Kramers problem: fifty years of development. *Phys. Rep.* **209**, 2–71 (1991).
- Mehta, A. D., Reif, M., Spudis, J. A., Smith, D. A. & Simmons, R. M. Single molecule biomechanics with optical methods. *Science* **283**, 1689–1695 (1999).
- Smith, D. E., Babcock, H. P. & Chu, S. Single-polymer dynamics in steady shear flow. *Science* **283**, 1724–1727 (1999).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312 (1997).
- Simon, A. & Libchaber, A. Escape and synchronization of a brownian particle. *Phys. Rev. Lett.* **68**, 3375–3378 (1992).
- Ghislain, L. P., Switz, N. A. & Webb, W. W. Measurement of small forces using an optical trap. *Rev. Sci. Instrum.* **65**, 2762–2768 (1994).

- Landauer, R. & Swanson, J. A. Frequency factors in the thermally activated process. *Phys. Rev.* **121**, 1668–1674 (1961).
- Risken, H. *The Fokker-Planck Equation* (Springer, Berlin, 1989).
- Hänggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).
- Linkwitz, S., Grabert, H., Turlot, E., Esteve, D. & Devoret, M. H. Escape rates in the region between the Kramers limits. *Phys. Rev. A* **45**, R3369–R3372 (1992).
- Luchinsky, D. G. & McClintock, P. V. E. Irreversibility of classical fluctuations studied in analogue electrical circuits. *Nature* **389**, 463–466 (1997).
- Smelyanskiy, V., Dykman, M. I. & Golding, B. Time oscillations of escape rates in periodically driven systems. *Phys. Rev. Lett.* **82**, 3193–3197 (1999).

Acknowledgements

We thank R. Kruse for his assistance with the experiment and graphics. Support from the Center for Fundamental Materials Research at Michigan State University and from the NSF Division of Physics, and Division of Materials Research is gratefully acknowledged.

Correspondence and requests for materials should be addressed to B.G. (email: golding@pa.msu.edu).

Injection and detection of a spin-polarized current in a light-emitting diode

R. Fiederling, M. Keim, G. Reuscher, W. Ossau, G. Schmidt, A. Waag & L. W. Molenkamp

Physikalisches Institut, EP III, Universität Würzburg, 97074 Würzburg, Germany

The field of magnetoelectronics has been growing in practical importance in recent years¹. For example, devices that harness electronic spin—such as giant-magnetoresistive sensors and magnetoresistive memory cells—are now appearing on the market². In contrast, magnetoelectronic devices based on spin-polarized transport in semiconductors are at a much earlier stage of development, largely because of the lack of an efficient means of injecting spin-polarized charge. Much work has focused on the use of ferromagnetic metallic contacts^{3,4}, but it has proved exceedingly difficult to demonstrate polarized spin injection. More recently, two groups^{5,6} have reported successful spin injection from an NiFe contact, but the observed effects of the spin-polarized transport were quite small (resistance changes of less than 1%). Here we describe a different approach, in which the magnetic semiconductor $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ is used as a spin aligner. We achieve injection efficiencies of 90% spin-polarized current into a non-magnetic semiconductor device. The device used in this case is a GaAs/AlGaAs light-emitting diode, and spin polarization is confirmed by the circular polarization state of the emitted light.

The quaternary II–VI magnetic semiconductor $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ has particular properties that make it ideally suitable as a spin-aligner for injecting electrons into GaAs. If its lattice parameter is kept constant and matched to the lattice constant of the GaAs substrate, then the Mn concentration can be varied over a wide range. The magnetic Mn ions are incorporated isoelectronically, and the conductivity type can be controlled by external dopants. This is in contrast to, for example, the III–V counterpart $\text{Ga}_x\text{Mn}_{1-x}\text{As}$, where the incorporation of Mn inherently leads to a high p-type doping⁷. Here we used n-type $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ as the spin-aligning material. Spin-injection of electrons rather than holes is advantageous because of the reduced spin–orbit coupling in the conduction band, decreasing spin decoherence. From the electrical point of view, $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ is also ideally suited as a spin-aligner on a GaAs-based heterostructure, as it allows for high quality

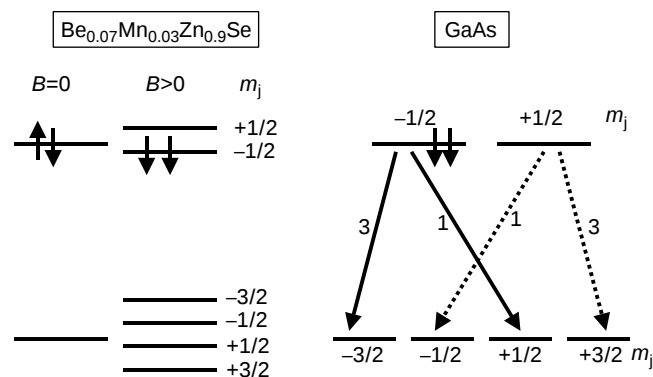


Figure 1 Conduction and valence band states of $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ and GaAs in an external magnetic field, B . The transition matrix element for heavy-hole transitions is a factor of 3 larger than for light-hole transitions.

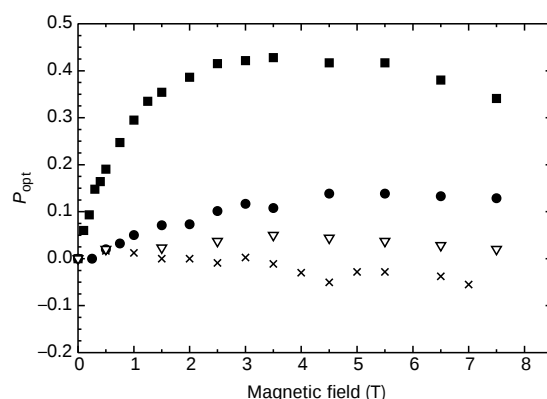


Figure 3 Degree of circular polarization of the electroluminescence. The results are shown as a function of the magnetic field strength for electron-injection across $\text{n-Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ spin-aligners ($d_{\text{SM}} = 300$ nm, squares), ($d_{\text{SM}} = 3$ nm, circles) and across a non-polarizing $\text{n-Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ electrode ($d_{\text{NM}} = 300$ nm, $d_{\text{SM}} = 0$ nm, triangles). The crosses represent the experimental data of the intrinsic polarization degree of the GaAs layer.

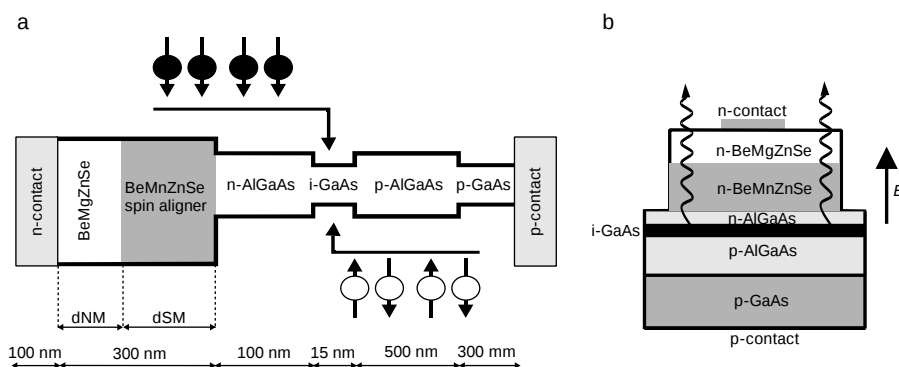


Figure 2 Device geometry and electric band structure. **a**, Schematic band structure of the spin-aligner light-emitting diode. Spin-polarized electrons are injected from the left into

the active GaAs layer, unpolarized holes from the right. **b**, Side view of the device showing the direction of the magnetic field and the emitted light.

interfaces with a conduction-band offset of less than 100 meV. Electrons may thus be injected at low energy, close to the bottom of the GaAs conduction band.

Manganese spins in undoped or n-type doped ZnSe based II–VI materials are usually anti-ferromagnetically coupled⁸. However, for low Mn-concentrations and at low temperatures the sp – d exchange interaction leads to a giant g -factor—the effective Landé factor of an electron inside the semiconductor—of up to 100, resulting in a large Zeeman splitting of band-edge-related states. (This is similar to the Zeeman splitting in EuSe which has been used previously as a spin-dependent tunnel barrier between a metal and a superconductor⁹.) We exploit the large Zeeman splitting to align the spins of all injected carriers to the energetically favourable lower Zeeman level. Subsequently these spin-polarized electrons are passed into the non-magnetic GaAs layer^{10,11}. Spin dephasing times are long for electrons in GaAs, so that the preferential spin orientation will survive for a diffusion length of up to 100 μm (refs 12, 13). Hence electric devices like light-emitting diodes (LEDs) or transistors can be driven with this spin-polarized current.

In order to detect the spin-polarized current in non-magnetic GaAs, it is convenient to study the degree of circular polarization of the electroluminescence emitted by an (Al,Ga)As diode¹⁴. In GaAs, the conduction band (s -character) is two-fold spin degenerate, whereas the valence band (p -character) is four-fold degenerate (heavy- and light-hole spin). The situation is schematically depicted

in Fig. 1. Spin degeneracy can be lifted in a magnetic field, and the light-hole / heavy-hole degeneracy can be lifted by, for example, biaxial strain or confinement. We neglect the small light-hole / heavy-hole splitting induced by the confinement in the 15-nm wide undoped GaAs layer, since the Al content of the $\text{Al}_{0.03}\text{Ga}_{0.97}\text{As}$ barriers used in our devices is small, and the quantum well width is large. When the conduction-band states are occupied with carriers of one spin orientation only, for example by injecting spin $+1/2$ electrons from a $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ contact, then according to the selection rule for the magnetic quantum number m_j ($\Delta m_j = \pm 1$), only two transitions are possible: one heavy-hole transition (from $+1/2$ to $+3/2$) and one light-hole transition (from $+1/2$ to $-1/2$). Whereas the heavy-hole transition is circularly polarized in one direction (called σ^+), the light-hole transition is polarized in the opposite direction (σ^-). The heavy-hole transition matrix element is a factor of 3 larger than that of the light-hole transition¹⁵, and therefore, the emitted light will be circularly polarized, with a degree of polarization of P_{opt} given by:

$$P_{\text{opt}} = \frac{(3n^{\uparrow} + n^{\downarrow}) - (3n^{\downarrow} + n^{\uparrow})}{(3n^{\uparrow} + n^{\downarrow}) + (3n^{\downarrow} + n^{\uparrow})}$$

Here $n^{\uparrow}$ and $n^{\downarrow}$ are the respective occupation numbers of the spin $\frac{1}{2}$ levels. This equation connects the circular polarization of the emitted electroluminescence with the polarization degree $P_j = (J^{\uparrow} - J^{\downarrow}) / (J^{\uparrow} + J^{\downarrow})$, where J is current density of the spin-polarized

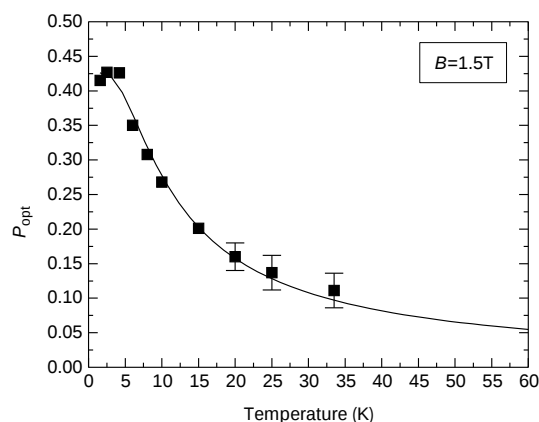


Figure 4 Temperature dependence of the degree of circular polarization of the electroluminescence of the $d_{\text{SM}} = 300$ nm diode. The line represents a fit using the splitting and occupation of the spin levels.

current. P_{opt} will depend on the original degree of spin-polarization of injected carriers, and the amount of depolarization due to scattering at the interface and during the diffusion into the quantum well.

We designed several relatively simple structures (Fig. 2) for demonstrating electrical spin-injection into GaAs. They consist of p-i-n diodes with a 15-nm-wide GaAs layer, embedded in $\text{Al}_{0.03}\text{Ga}_{0.97}\text{As}$ barriers. The top n-contacts consist of the semimagnetic $\text{Be}_{0.07}\text{Mn}_{0.03}\text{Zn}_{0.9}\text{Se}$ with thickness d_{SM} and/or a non-magnetic BeMgZnSe layer of thickness d_{NM} . The total thickness of the II–VI top layer ($d_{\text{SM}} + d_{\text{NM}} = 300$ nm) is kept constant in order to maintain the same overall injection conditions. A total of four wafers was grown, and several devices were fabricated from each wafer. All devices exhibit consistent behaviour.

When a small external magnetic field is applied, the electrons passing through the II–VI top layer will become spin-polarized in the $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ layer. The polarization degree of the injected electrons will depend on the thickness of the semimagnetic layer d_{SM} . The holes are injected from the bottom p-GaAs substrate and pass through p- $\text{Al}_{0.03}\text{Ga}_{0.97}\text{As}$; hence they are unpolarized. Both p- $\text{Al}_{0.03}\text{Ga}_{0.97}\text{As}$ and n-doped II–VI layers are heavily doped in order to prevent freeze-out (rebinding of electrons and holes at the dopant atoms) of carriers at low temperatures. In order to minimize the dephasing and scattering of the injected carriers' spin, the distance between the injecting $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ layer and the GaAs layer has been kept small (100 nm). In addition, the $\text{Al}_{0.03}\text{Ga}_{0.97}\text{As}$ spacer in-between was only slightly doped (10^{16} cm^{-3}). It has been shown that the preservation of electron spin over long distances is maximized at such a doping level¹⁶.

In Fig. 3, we show P_{opt} of the electroluminescence under forward bias as a function of the magnetic field. We indeed detect a strongly polarized electroluminescence, with a maximum P_{opt} of 43% for the structure with the thick semimagnetic injection layer (squares in Fig. 3). P_{opt} increases with magnetic field until it saturates for fields above 3 T. This behaviour reflects the degree of spin-polarization of the injected carriers achieved in the $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ spin-aligner, where the Zeeman splitting follows a typical Brillouin function, and saturates when all carriers are aligned along the direction of the external magnetic field. Taking into account the four-fold degeneracy of the valence band, this corresponds to an electron spin-polarization P_j of almost 90%. The temperature dependence of P_{opt} has been studied and is shown in Fig. 4. The value of the g-factor in $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ decreases drastically with increasing temperature⁸, according to the splitting and occupation of the spin-levels. Again, this behaviour is reflected by the polarization of the electrolumi-

nescence, which decreases with temperature and follows the correct functional dependence (a hyperbolic tangent with a Brillouin function as argument).

In order to separate the effects of spin-injection from other effects that could also lead to polarized luminescence, we made special efforts to ensure that polarization effects of the monochromator grating, mirrors, windows and the excitation source were eliminated. The bandgap of the $\text{Be}_x\text{Mn}_y\text{Zn}_{1-x-y}\text{Se}$ spin-aligner is twice as large as that of the active GaAs layer, avoiding possible complications from field-induced optical dichroism. We verified that the handedness of the circular optical polarization changes with magnetic field reversal. In addition, we fabricated structures with magnetic injectors of reduced thickness $d_{\text{SM}} = 3$ nm as well as structures with a fully non-magnetic injector ($d_{\text{NM}} = 300$ nm). The structures with $d_{\text{SM}} = 3$ nm show a reduced degree of optical polarization of 11% at high magnetic fields (circles in Fig. 3) indicating that the layer is too thin to achieve complete relaxation of all spins into the lower Zeeman level. The structure with a non-magnetic injector ($\text{Be}_x\text{Mg}_y\text{Zn}_{1-x-y}\text{Se}$) shows almost no polarization of the electroluminescence (triangles in Fig. 3). We have also measured the degree of optical polarization of the emission from the GaAs layer (crosses in Fig. 3) obtained by photoexcitation (using unpolarized light) in the barrier layers. The results of this experiment rule out any contribution of field-induced optical dichroism in the electroluminescence of the diode: the degree of polarization in the photoluminescence is very small and opposite in sign to that obtained for the experiment with spin-polarized electron injection. This behaviour can be fully attributed to the intrinsic g-factors of the GaAs layer. In the electroluminescence of the diodes, the intrinsic polarization of the GaAs layer only becomes important when the polarization degree of the spin-injected electrons saturates. This is the reason for the observed decrease of P_{opt} for the electroluminescence at high magnetic fields (squares in Fig. 3).

These combined results can only be interpreted as clear evidence that the polarized electroluminescence of the spin-aligner LED is due to the injection of spin-polarized carriers into GaAs. The very large degree of spin-polarization achieved in this device clearly illustrates that the spin-aligner approach, although limited to low temperatures, seems ideally suited to test the many outstanding predictions on the physics of spin transport. In this context, the moderate magnetic fields needed to obtain a sizeable Zeeman splitting in the spin-aligner layer can conveniently be obtained by using a ferromagnetic metallic contact.

Received 5 July; accepted 19 October 1999.

1. Simmonds, J. L. Magneto-electronics today and tomorrow. *Phys. Today* **48**(4), 26–32 (1995).
2. Parkin, S. S. P. *et al.* Exchange-biased magnetic tunnel junctions and application to nonvolatile magnetic random access memory. *J. Appl. Phys.* **85**, 5828–5833 (1999).
3. Datta, S. & Das, B. Electronic analog of the electro-optic modulator. *Appl. Phys. Lett.* **56**(7), 665–667 (1990).
4. Prinz, G. A. Spin-polarized transport. *Phys. Today* **48**(4), 58–63 (1995).
5. Hammar, P. R., Bennet, B. R., Yang, M. J. & Johnson, M. Observation of spin injection at a ferromagnet-semiconductor interface. *Phys. Rev. Lett.* **83**(1), 203–206 (1999).
6. Gardelis, S., Smith, C. G., Barnes, C. H. W., Linfield, E. H. & Ritchie, D. A. Spin-value effects in a semiconductor field effect transistor: A spintronic device. *Phys. Rev. B* **60**(11), 7764–7767 (1999).
7. Matsukura, F., Ohno, H., Shen, A. & Sugawara, Y. Transport properties and origin of ferromagnetism in (Ga,Mn)As. *Phys. Rev. B* **57**, R2037–R2040 (1998).
8. Furdyna, J. K. Diluted magnetic semiconductors. *J. Appl. Phys.* **64**, R29–R64 (1988).
9. Moodera, J. S., Meserve, R. & Hao, X. Variation of the electron-spin polarization in EuSe tunnel junctions from zero to near 100% in a magnetic field. *Phys. Rev. Lett.* **70**(6), 853–856 (1993).
10. Oestreich, M. *et al.* Spin injection into semiconductors. *Appl. Phys. Lett.* **74**, 1251–1253 (1999).
11. Carlos Eguia J. Spin-dependent perpendicular magnetotransport through a tunable $\text{ZnSe}/\text{Zn}_{1-x}\text{Mn}_x\text{Se}$ heterostructure: A possible spin filter. *Phys. Rev. Lett.* **80**(20), 4578–4581 (1998).
12. Hägele, D., Oestreich, M., Rühle, W. W., Nestle, N. & Eberl, K. Spin transport in GaAs. *Appl. Phys. Lett.* **73**(1), 1580–1582 (1998).
13. Kikkawa, J. M. & Awschalom, D. D. Lateral drag of spin coherence in gallium arsenide. *Nature* **397**, 139–141 (1999).
14. Aronov, A. G. & Pikus, G. E. Spin injection into semiconductors. *Sov. Phys. Semicond.* **10**, 698–700 (1976).
15. Weisbuch, C. & Vinter, B. *Quantum Semiconductor Structures—Fundamentals and Applications* (Academic, Boston, 1991).
16. Awschalom, D. D. & Kikkawa, J. M. Electron spin and optical coherence in semiconductors. *Phys. Today* **6**, 33–38 (1999).

Acknowledgements

We acknowledge the financial support of the European Commission (Esprit project SPIDER) as well as the Bundesministerium für Bildung und Forschung (Verbundprojekt Elektronische Korrelationen und Magnetismus). We would also like to thank Th. Gruber, V. Hock, B. König, D. Yakovlev, G. Müller and G. E. W. Bauer for discussions and help.

Correspondence and requests for materials should be addressed to L.W.M. (e-mail: laurens.molenkamp@physik.uni-wuerzburg.de).

Electrical spin injection in a ferromagnetic semiconductor heterostructure

Y. Ohno*, D. K. Young†, B. Beschoten†, F. Matsukura*, H. Ohno* & D. D. Awschalom†

* Laboratory for Electronic Intelligent Systems, Research Institute of Electrical Communication, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan

† Center for Spintronics and Quantum Computation, Quantum Institute, University of California, Santa Barbara, California 93106, USA

Conventional electronics is based on the manipulation of electronic charge. An intriguing alternative is the field of 'spintronics', wherein the classical manipulation of electronic spin in semiconductor devices gives rise to the possibility of reading and writing non-volatile information through magnetism^{1,2}. Moreover, the ability to preserve coherent spin states in conventional semiconductors³ and quantum dots⁴ may eventually enable quantum computing in the solid state^{5,6}. Recent studies have shown that optically excited electron spins can retain their coherence over distances exceeding 100 micrometres (ref. 7). But to inject spin-polarized carriers electrically remains a formidable challenge^{8,9}. Here we report the fabrication of all-semiconductor, light-emitting spintronic devices using III–V heterostructures based on gallium arsenide. Electrical spin injection into a non-magnetic semiconductor is achieved (in zero magnetic field) using a p-type ferromagnetic semiconductor¹⁰ as the spin polarizer. Spin polarization of the injected holes is determined directly from the polarization of the emitted electroluminescence following the recombination of the holes with the injected (unpolarized) electrons.

The device structure shown in Fig. 1a was grown by molecular beam epitaxy on a (100) n-GaAs substrate with a 500 nm n⁺-GaAs buffer layer (doping density $N_D = 2 \times 10^{18} \text{ cm}^{-3}$) and the following layers: a 20-nm-thick undoped GaAs layer, a 10-nm-thick undoped $\text{In}_{0.13}\text{Ga}_{0.87}\text{As}$ strained quantum well (QW), an undoped GaAs spacer with thickness d (20, 120, 140 or 220 nm), and a 300-nm-thick $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ layer with $x = 0.045$. GaAs and (In,Ga)As layers were grown at $T = 540\text{--}580^\circ\text{C}$, while the (Ga,Mn)As layer was grown at $T = 250^\circ\text{C}$. The 4.5% Mn concentration is determined from the lattice constant measured by X-ray diffraction¹¹, and is expected to yield the Curie temperature T_C in the range of 40–90 K with a hole concentration $p \sim 10^{20} \text{ cm}^{-3}$ (ref. 12). The easy axis of the (Ga,Mn)As magnetization is in the plane of the sample, verified by a superconducting quantum interference device (SQUID) magnetometer¹¹. The (In,Ga)As QW allows us to study the depth of spin injection¹³, when employed as a "spin detector" placed at various distances from the (Ga,Mn)As layer. The epitaxial wafer was processed into light-emitting devices having 200- μm -wide mesa stripes defined by wet chemical etching after metal

electrode deposition (a 5-nm-thick chromium film and a 150-nm-thick gold film) and cleaved into 1-mm pieces. Two sets of control samples were prepared to verify spin injection, including a non-magnetic device ($d = 20 \text{ nm}$) with a p-type GaAs:Be layer ($N_D = 2 \times 10^{18} \text{ cm}^{-3}$) substituted for the (Ga,Mn)As layer and a magnetic structure ($d = 100 \text{ nm}$) without metal contacts enabling optical excitation of the QW.

Spontaneous magnetization develops below T_C in (Ga,Mn)As, creating spin-polarized holes¹⁴. Under forward-bias conditions, spin-polarized holes are injected into the QW through the undoped GaAs spacer layer (Fig. 1a), while unpolarized electrons are supplied from the bottom n-GaAs substrate. The devices were driven with a 1-kHz sinusoid to minimize sample heating. A representative current–voltage (I – V) plot characteristic of the devices, with $d = 20 \text{ nm}$, is shown in the inset of Fig. 1b. The samples were mounted in a magneto-optical cryostat with a variable magnetic field that was monitored by *in situ* Hall bars, and applied in the easy plane of the (Ga,Mn)As layer. Electroluminescence (EL) was collected from the cleaved facet to minimize magneto-optical effects due to the nearby (Ga,Mn)As. The polarization $P = (\sigma^+ - \sigma^-)/(\sigma^+ + \sigma^-)$ of the EL spectra was analysed with a variable wave plate and linear polarizer, and detected with a charge-coupled device attached to a 0.5-m spectrometer. Here σ^+ and σ^- are the energy-integrated (shaded grey area in Fig. 1b) intensity of right and left circularly polarized components of the EL peak respectively. Figure 1b shows the spectrally resolved total EL intensity (black curve) and polarization (red curve) at temperature $T = 6 \text{ K}$ and magnetic field $H = 1000 \text{ Oe}$ of a device with $d = 20 \text{ nm}$. A peak in the EL intensity (full-width at half-maximum, FWHM = 13 meV)

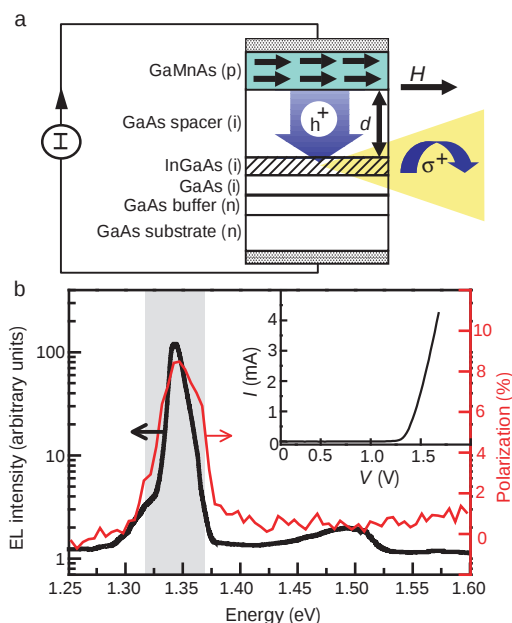


Figure 1 Electrical spin injection in an epitaxially grown ferromagnetic semiconductor heterostructure, based on GaAs. **a**, Spontaneous magnetization develops below the Curie temperature T_C in the ferromagnetic p-type semiconductor (Ga,Mn)As, depicted by the black arrows in the green layer. Under forward bias, spin-polarized holes from (Ga,Mn)As and unpolarized electrons from the n-type GaAs substrate are injected into the (In,Ga)As quantum well (QW, hatched region), through a spacer layer with thickness d , producing polarized EL. **b**, Total electroluminescence (EL) intensity of the device ($d = 20 \text{ nm}$) under forward bias at temperature $T = 6 \text{ K}$ and magnetic field $H = 1,000 \text{ Oe}$ is shown (black curve) with its corresponding polarization (red curve). Current $I = 1.43 \text{ mA}$. Note that the polarization is largest at the QW ground state ($E = 1.34 \text{ eV}$). The EL and polarization are plotted on semi-log and linear scales, respectively. Inset, a current–voltage plot characteristic of a 20-nm spacer layer device. Shaded grey area, see Fig. 2.

and polarization is seen at the QW ground-state transition energy ($E = 1.34$ eV). The measured polarization contains a constant background due most likely to a combination of sample strain and experimental geometry. We note that the polarization amplitude is sensitive to the angle of collection, which may reflect variation of the effective field or directionality in the polarization induced by the waveguide of the structure. We also observe no systematic dependence of the polarization on either the spacer layer thickness or the injection current density.

Figure 2 shows relative changes in EL polarization, $\Delta P \equiv P - P_{\text{bkgd}}$, where P_{bkgd} is the background polarization, as a function of in-plane magnetic field for various temperatures above and below T_C . Below $T = 31$ K, ΔP is hysteretic with loops closing at $H \leq 300$ Oe. The coercive field increases from 0 to 40 Oe as the temperature decreases from T_C to $T = 6$ K. The inset shows that the temperature dependence of the relative remanent polarization (ΔP at $H = 0$ Oe) at $T = 6$ –94 K follows the sample's magnetic moment M (solid black curve), independently measured using a SQUID magnetometer. The deviation from Brillouin behaviour of $M(T)$ is consistent with studies of the magnetization temperature dependence in (Ga,Mn)As^{14,15}. No polarization hysteresis is observed for $T \geq 52$ K, when (Ga,Mn)As becomes paramagnetic.

A non-magnetic device was grown in order to verify that the hysteresis in the polarization is due to spin injection rather than carrier polarization from the (Ga,Mn)As stray field. In contrast to the magnetic devices, the EL polarization from this non-magnetic device (Fig. 3a) shows no hysteresis at $T = 6$ K from -500 Oe $< H < 500$ Oe, thus linking the hysteretic polarization to the presence of a (Ga,Mn)As layer. Furthermore, the field range over which we observe no change in the polarization is comparable to the stray field expected at the (Ga,Mn)As facet (~ 500 Oe)¹¹. This suggests that local fields from the (Ga,Mn)As layer are unlikely to be responsible for the hysteretic polarization in the magnetic structures.

Since (Ga,Mn)As exhibits strong magnetic circular dichroism¹⁴, it was also important to confirm that the observed hysteretic polarization is not due to possible re-absorption of QW luminescence passing through the (Ga,Mn)As layer. A magnetic sample without metal contacts was prepared, allowing direct optical excitation of

unpolarized carriers into the QW. A p-type layer between the QW and a semi-insulating substrate was incorporated into the structure in order to reduce the electrostatic potential across the junction, thus leading to more efficient radiative recombination. Selective spectral excitation (FWHM ~ 1 meV) by a continuous-wave Ti-sapphire laser was used to create unpolarized carriers in the QW by illuminating through the polished substrate (~ 80 μm) with linearly polarized light at $E = 1.398$ eV, which is 53 meV above the QW ground state. Light was collected from the cleaved facet with an identical optical path to the EL measurements. The photoluminescence (PL) polarization as a function of magnetic field, shown in Fig. 3b, shows no hysteresis, providing further evidence that spin injection is occurring in the magnetic samples.

The presence of hysteretic polarization observed in magnetic samples with $d = 20$ –220 nm, and its absence in the control samples, indicate that hole spins can be injected and transported over 200 nm. This occurs in spite of possible depolarization due to transport through interfaces, as well as valence band mixing in the bulk and QW regions. Additionally, hole spins can be preserved during energy relaxation in a QW¹⁶. As opposed to the high energy injection of "hot" carriers from metal/semiconductor junctions, energy relaxation from semiconductor p–n junctions that are in quasi-equilibrium is minimized, which may determine the length scale of spin transport in the present structure. It is still difficult to estimate the expected degree of EL polarization from these types of structures, as detailed knowledge about the relevant band structure and transport/optical properties of (Ga,Mn)As is not yet available.

We have used epitaxially grown all-semiconductor ferromagnetic/non-magnetic structures to demonstrate electrical spin injection and transport. Electroluminescence polarization reveals that the hole spin polarization in a ferromagnetic semiconductor can be transported across interfaces over distances greater than 200 nm in a forward biased p–n junction structure, and switched in modest fields of $H \sim 40$ Oe. The demonstration of electrical spin injection into a semiconductor, especially in the absence of an applied magnetic field, suggests a variety of new technological opportunities, including electrical control of magnetization, integration of nonvolatile storage and logic, and manipulation of quantum spin states in semiconductors. $\square$

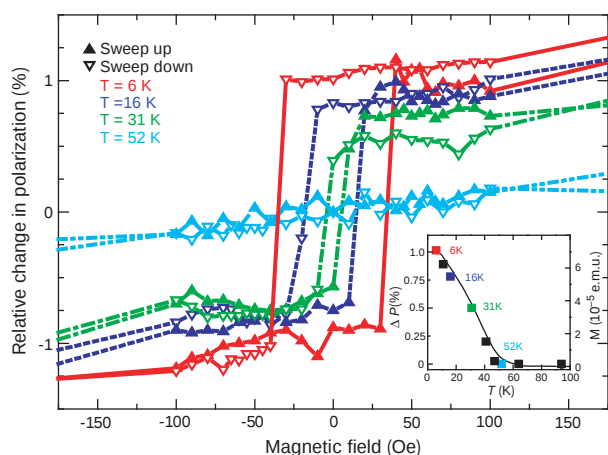


Figure 2 Hysteretic electroluminescence polarization is a direct result of spin injection from the ferromagnetic (Ga,Mn)As layer. Shown are relative changes in the energy-integrated (shaded grey area in Fig. 1b) polarization ΔP , at temperatures $T = 6$ –52 K, as a function of in-plane field from a device with $d = 140$ nm. $E = 1.34$ eV, $I = 2.8$ mA. Triangles indicate points taken when the field is swept up or down. Inset, the relative remanent polarization (ΔP at $H = 0$ Oe) shown in solid squares at $T = 6$ –94 K, and the temperature dependence of the (Ga,Mn)As magnetic moment, measured by a SQUID magnetometer (solid black curve), demonstrating that polarization is proportional to magnetic moment.

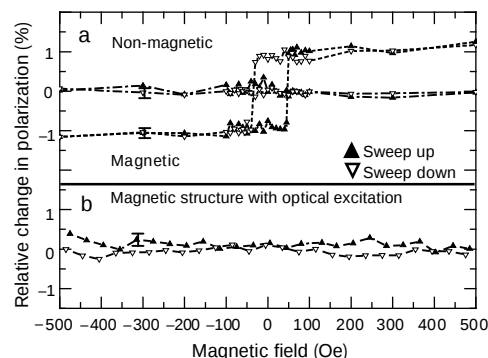


Figure 3 The absence of hysteretic polarization. Data shown from a forward biased non-magnetic device at $T = 6$ K (a) and from a magnetic structure under optical excitation (b) verify that electrical spin injection occurs within the device depicted in Fig. 1a. a, Magnetic field dependence of ΔP from a non-magnetic device is shown in contrast to a magnetic structure, with identical layers ($d = 20$ nm). Triangles as for Fig. 2. b, No hysteresis is observed from the photoluminescence (PL) polarization of a magnetic sample excited with $E = 1.398$ eV (QW PL ground state, $E = 1.345$ eV).

Received 9 September; accepted 5 November 1999.

1. Awschalom, D. D. & Kikkawa, J. M. Electron spin and optical coherence in semiconductors. *Phys. Today* **52**, 33–38 (1999).
2. Prinz, G. A. Magneto-electronics. *Science* **282**, 1660–1663 (1998).
3. Kikkawa, J. M. & Awschalom, D. D. Resonant spin amplification in n-type GaAs. *Phys. Rev. Lett.* **80**, 4313–4316 (1998).
4. Gupta, J. A., Awschalom, D. D., Peng, X. & Alivisatos, A. P. Spin coherence in semiconductor quantum dots. *Phys. Rev. B* **59**, 10421–10424 (1999).
5. DiVincenzo, D. P. Quantum computation. *Science* **270**, 255–261 (1995).
6. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
7. Kikkawa, J. M. & Awschalom, D. D. Lateral drag of spin coherence in gallium arsenide. *Nature* **397**, 139–141 (1999).
8. Monzon, F. G. & Roukes, M. L. Spin injection and the local hall effect in InAs quantum wells. *J. Mag. Magn. Mater.* **198**, 632–635 (1999).
9. Hammar, P. R. *et al.* Observation of the spin injection at a ferromagnetic-semiconductor interface. *Phys. Rev. Lett.* **83**, 203–206 (1999).
10. Ohno, H. Making nonmagnetic semiconductors ferromagnetic. *Science* **281**, 951–955 (1998).
11. Ohno, H. *et al.* (Ga,Mn)As: A new diluted magnetic semiconductor based on GaAs. *Appl. Phys. Lett.* **69**, 363–365 (1996).
12. Matsukara, F., Ohno, H., Shen, A. & Sugawara, Y. Transport properties and origin of ferromagnetism in (Ga,Mn)As. *Phys. Rev. B* **57**, 2037–2040 (1998).
13. Hägele, D., Oestreich, M., Rühle, W. W., Nestle, N. & Eberl, K. Spin transport in GaAs. *Appl. Phys. Lett.* **73**, 1580–1582 (1998).
14. Beschoten, B. *et al.* Magnetic circular dichroism studies of carrier-induced ferromagnetism in Ga_{1-x}Mn_xAs. *Phys. Rev. Lett.* **83**, 3073–3076 (1999).
15. Van Esch, A. *et al.* Interplay between the magnetic and transport properties in the III-V diluted magnetic semiconductor Ga_{1-x}Mn_xAs. *Phys. Rev. B* **56**, 13103–13112 (1997).
16. Uenoyama, T. & Sham, L. J. Hole relaxation and luminescence polarization in doped and undoped quantum wells. *Phys. Rev. Lett.* **64**, 3070–3073 (1990).

Acknowledgements

We thank I. Arata for technical support and D. T. Fuchs and J. M. Kikkawa for critical readings of the manuscript. Work at UCSB is supported by the Air Force Office of Scientific Research, the National Science Foundation through the Center for Quantized Electronic Structures, and the Office of Naval Research. The Japan Society for the Promotion of Science and the Ministry of Education in Japan support work at Tohoku University.

Correspondence and requests for materials should be addressed to D.D.A. (e-mail: awsch@physics.ucsb.edu).

Lithium-doped plastic crystal electrolytes exhibiting fast ion conduction for secondary batteries

Douglas R. MacFarlane*, Junhua Huang* & Maria Forsyth†

* Department of Chemistry, † Department of Materials Engineering, Monash University, Clayton, Victoria 3168, Australia

Rechargeable lithium batteries have long been considered an attractive alternative power source for a wide variety of applications. Safety and stability¹ concerns associated with solvent-based electrolytes has necessitated the use of lithium intercalation materials (rather than lithium metal) as anodes, which decreases the energy storage capacity per unit mass. The use of solid lithium ion conductors—based on glasses, ceramics or polymers—as the electrolyte would potentially improve the stability of a lithium-metal anode while alleviating the safety concerns. Glasses and ceramics conduct via a fast ion mechanism, in which the lithium ions move within an essentially static framework. In contrast, the motion of ions in polymer systems is similar to that in solvent-based electrolytes—motion is mediated by the dynamics of the host polymer, thereby restricting the conductivity to relatively low values. Moreover, in the polymer systems, the motion of the lithium ions provides only a small fraction of the overall conductivity², which results in severe concentration gradients

during cell operation, causing premature failure³. Here we describe a class of materials, prepared by doping lithium ions into a plastic crystalline matrix, that exhibit fast lithium ion motion due to rotational disorder and the existence of vacancies in the lattice. The combination of possible structural variations of the plastic crystal matrix and conductivities as high as $2 \times 10^{-4} \text{ S cm}^{-1}$ at 60 °C make these materials very attractive for secondary battery applications.

A fast ion conductor is a substance in which one of the ionic species is able to conduct much more rapidly—often by a factor of many orders of magnitude—than any other species present. A typical example is the ceramic material $\text{Li}_{1-x}\text{Al}_x\text{Ti}_{2-x}(\text{PO}_4)_3$ (refs 4, 5); at room temperature the ions involved in the structure of this material are essentially immobile, in the sense of long-range diffusive and conductive motions. In contrast, the Li^+ ions exhibit conductivity as high as $10^{-3} \text{ S cm}^{-1}$ at room temperature. The occurrence of this decoupling of the motion of one ion from the motions of others has been of fundamental interest for many years. Models of fast ion conduction in ceramics consider the motion to be along interconnected tunnels in the solid lattice and/or involve ion hopping between empty sites in the lattice. Although the single-crystal conductivity of these materials can be high, the polycrystalline nature of the bulk material creates grain boundaries which lower the overall conductivity.

Plastic crystalline substances have also been known for many years⁶ and those based on ionic species can exhibit ion conduction^{7–15}. The crystalline state of these substances is typically fully ordered at low temperatures, but as the temperature is increased one or more first-order solid–solid phase transitions are observed. In these higher-temperature phases, there exists some degree of rotational disorder; for example, in Li_2SO_4 (refs 13–15) the SO_4^{2-} anion is free to rotate in the face-centred cubic (f.c.c.) phase of this compound above 575 °C, but is nonetheless fixed on its lattice site. Also, organic ion plastic crystals have been studied^{8–12}, and most exhibit some level of intrinsic conductivity. This is understood to be associated with the rotational motions of one or more of the ions; motion of the lattice defects is also involved. ‘Double salt’ compounds have been prepared¹² that showed plastic crystal behaviour, and had conductivities as high as $4 \times 10^{-6} \text{ S cm}^{-1}$ at 25 °C, the bulk of which was attributed to Li^+ ion motion. A new family of plastic crystal ionic compounds, involving the alkylmethylpyrrolidinium cation (I in Fig. 1) and the bis(trifluoromethanesulphonyl)imide anion (II in Fig. 1, referred to here as the ‘imide’ ion) was recently reported^{16,17}. The salts are referred to via an acronym, P_{1x} , the subscripts indicating the number of carbons in each alkyl chain. They show a characteristic series of solid–solid phase transitions below their melting points, and have ambient-temperature conductivity between 10^{-7} and $10^{-9} \text{ S cm}^{-1}$.

Here we obtained phases with fast lithium ion conduction by doping a lithium compound into these P_{1x} phases. The organic salt acts as a solid-state solvent or ‘matrix’ for the Li ions. The lithium salt contains the same anion as the matrix phase, hence the doping can be considered as a cation substitution. Doping is achieved by

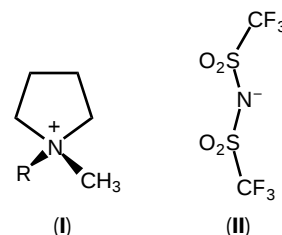


Figure 1 Structure of the alkylmethylpyrrolidinium imide matrix compounds. In the text, II is referred to as ‘imide’.

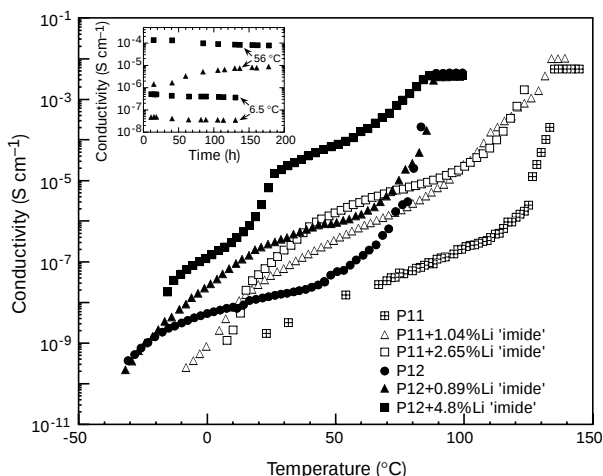


Figure 2 Ionic conductivity as a function of temperature for P_{11} and P_{12} solid phases doped with Li^+ . Inset, conductivity versus time, at temperatures shown, indicating that all samples approach a conductivity limit. In the case of P_{12} doped with 0.89 mol% Li 'imide', the conductivity at 56 °C in fact increases. This may represent a homogenizing of the dopant in the sample. The fact that little change is seen in conductivity over a period of more than 1 week indicates that the phases are stable at least over this time period. The rate of diffusion-controlled events can be expected to be slower at the lower temperature.

mixing the lithium salt (3M) into the molten state of the P_{1x} compound, and then cooling the material into its plastic crystalline state. The result is a material that is a waxy solid at room temperature in the cases of the P_{11} and P_{12} members of the family. Electrochemical studies show that Li(s) can be reduced from the doped materials, and that they are stable over a potential 'window' of ~ 5 V.

The conductivity behaviour of these doped P_{1x} phases over a wide range of temperatures—measured by standard impedance spectroscopy methods^{16,17} and after thorough drying of the material—is unexpectedly complex (Fig. 2). As noted above, the pure matrix phase compounds (P_{11} and P_{12}) are slightly conductive, their conductivity rising sharply at their final melting points. Doping with less than 1 cation% Li^+ produces a 20-fold increase in conductivity at 25 °C in both cases. Figure 3 presents the effect of composition on conductivity up to 50 mol% Li 'imide'. Beyond this point the conductivity continues to fall; pure Li 'imide' is practically non-conductive. Whereas at low compositions only one phase is

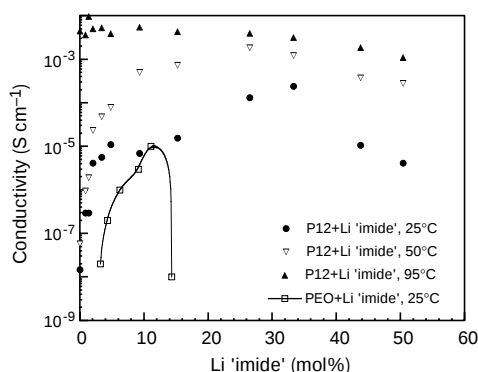


Figure 3 Dependence of conductivity on Li content in the P_{12} -Li 'imide' binary system. The PEO-Li 'imide' data from ref. 18 refers to a polyethylene oxide based polymer electrolyte material.

observed and Li^+ can be considered to be a dopant in the matrix phase, at concentrations between 5 and 45 mol% Li 'imide' we observed a eutectic transition in the thermal data at 30 °C indicating the presence of two phases in the samples. Thus the 25 °C isotherm in Fig. 3 represents solid-state conductivity in all cases (as discussed further below). The 95 °C isotherm corresponds to liquid-state samples in all cases; in contrast to the doped solid-state materials, the addition of small amounts of Li ions to this molten state produces a gradual decrease in conductivity. The comparison in Fig. 3 with the Li 'imide' containing polymer electrolyte¹⁸ indicates that the plastic crystal materials are as conductive as analogous solid polymer electrolytes, but at substantially lower lithium ion concentrations. On a gravimetric basis the highest conductivity in the polymer electrolyte case is reached beyond 1 mol kg⁻¹ at 25 °C (ref. 18), whereas equivalent conductivity is reached in the plastic crystal materials at less than 0.1 mol kg⁻¹.

Some light is shed on the complexity of the conductivity behaviour by the thermograms shown in Fig. 4, which were obtained by thermal analysis of the samples at a heating rate of 10 °C min⁻¹. The thermograms show a number of solid-solid phase transitions that are present in both the pure host (P_{12}) and in the doped materials ($P_{12} + 0.89\%$ Li). Some of these phase transitions, the first of which takes place at -90 °C, correspond to the onset of rotational motion of one or more parts of the molecules involved^{7-11,16}. Whether it is the cation or the anion in the matrix phase that is responsible for the rotational disorder is not yet known. The melting transition of the doped compound is broadened and slightly shifted compared to the pure compound. In addition, no new thermal transitions—for example, eutectic melting—are observed, indicating that the material probably represents a solid solution of the dopant in the parent structure. (The thermal behaviour of P_{11} -based systems is similar, showing solid-solid transitions at -90 and 81 °C before melting at 132 °C.) At much higher Li 'imide' concentrations the P_{12} -based material exhibits a clear eutectic transition as well as the final melting (data not shown). It is therefore likely that the behaviour of the 4.8 cation% sample in Fig. 2 is a result of it being a multiphase mixture; nonetheless, the material retains its waxy solid properties close to its final melting point. At 50 mol% Li 'imide' the thermal data (not shown) suggest that a new, possibly metastable, mixed salt compounds forms, as observed in a study of double salts¹³.

That there are diffusive motions of the ions present in these materials at temperatures well below the melting point has been confirmed by NMR experiments (Fig. 5) in which the width of the ^7Li resonance is probed as a function of temperature. The ^7Li linewidth tends towards a low-temperature limit, often referred to as the rigid lattice linewidth, characteristic of lithium ions trapped

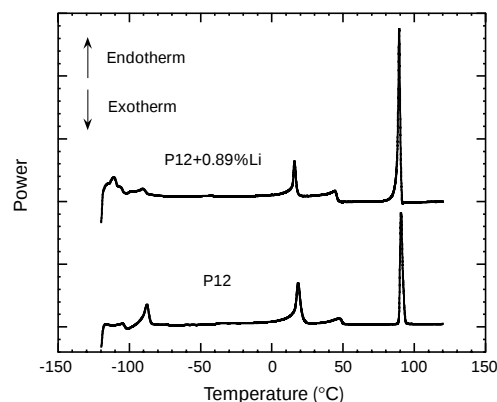


Figure 4 Thermal analysis traces obtained by differential scanning calorimetry of pure and Li^+ -doped P_{12} .

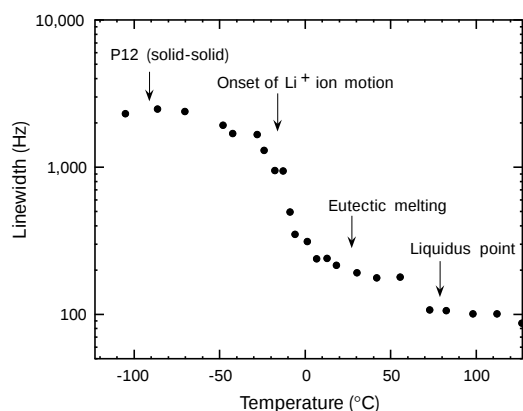


Figure 5 ^7Li NMR linewidth data as a function of temperature for P_{12} doped with 9.3 mol% Li^+ .

in a non-diffusing state in a static matrix. With increasing temperature, the onset of diffusive motions causes the linewidth to fall towards values characteristic of a fluid liquid. In a simple lithium compound, this line narrowing would take place just below the melting point. In contrast, two distinct steps as a function of increasing temperature are observed in Fig. 5. The higher-temperature step corresponds to the melting transition in P_{12} . The much larger line-narrowing event takes place around -20°C and is complete by 0°C , well below the melting point of the compound. The diffusive motions causing line narrowing could involve the lithium ion itself or its immediate neighbours. Line narrowing therefore does not offer conclusive proof of lithium ion motion below the melting point. However, in combination with the observation that the addition of the lithium ion dopant, along with an anion which is common to the matrix, produces a 50-fold increase in conductivity without altering the phase behaviour of the material, it appears likely that it is lithium ion motion which is responsible for both the NMR line narrowing and the increased conductivity. Conduction thus being substantially due to the Li^+ ion alone is in stark contrast to polymer electrolyte materials in which lithium ion motion makes up only a small ($<20\%$) fraction of the ion conductivity².

The doped P_{1x} materials we report here can be classified as fast ion conductors on the basis of the high conductivities observed in a crystalline lattice. These crystalline materials have plastic mechanical properties at room temperature; that is, at some value of applied stress the materials will yield and exhibit plastic flow. Both the flow and the fast ion conduction properties are a result of rotational disorder and the existence of vacancies in the lattice. The materials therefore express an important decoupling of various molecular motions from one another such that an unusual combination of properties is exhibited. NMR relaxation measurements in progress should determine the various motional timescales that are present in these materials. □

Received 10 May; accepted 4 November 1999.

- Gray, F. M. *Solid Polymer Electrolytes: Fundamentals and Technological Applications* 13 (VCH, New York, 1991).
- Ma, Y. et al. The measurement of a complete set of transport properties for a concentrated solid polymer electrolyte solution. *J. Electrochem. Soc.* **142**, 1859–1868 (1995).
- Doeff, M. M., Ferry, A., Ma, Y., Ding, L. & De Jonghe, L. C. Effect of electrolyte composition on the performance of sodium/polymer cells. *J. Electrochem. Soc.* **144**, L20–L22 (1997).
- Wong, S. et al. Towards elucidating microscopic structural changes in Li-ion conductors $\text{Li}_{1-x}\text{Ti}_x\text{Al}_2(\text{PO}_4)_3$ and $\text{Li}_{1-x}\text{Ti}_x\text{Al}_2(\text{PO}_4)_3-x(\text{MO}_4)_x$ ($M = \text{V}$ and Nb): x-ray and ^{27}Al and ^{31}P NMR studies. *J. Mater. Chem.* **8**, 2199–2203 (1998).
- Adachi, G.-Y., Imanaka, N. & Aono, H. Fast Li^+ conducting ceramic electrolytes. *Adv. Mater.* **8**, 127–134 (1996).
- Timmermans, J. Plastic crystals: a historical review. *J. Phys. Chem. Solids* **18**, 1–8 (1961).
- Handra, D., Helmes, J. H. & Majumdar, A. Ionic conductivity in ordered and disordered phases of plastic crystals. *J. Electrochem. Soc.* **141**, 1921–1927 (1994).

- Hattori, M., Fukada, S., Nakamura, D. & Ikeda, R. Studies of the anisotropic self-diffusion and reorientation of butyl ammonium cations in the rotator phase of butyl ammonium chloride using ^1H magnetic resonance, electrical conductivity and thermal measurements. *J. Chem. Soc. Faraday Trans.* **86**, 3777–3783 (1990).
- Ishida, H., Furukawa, Y., Kashino, S., Sato, S. & Ikeda, R. Phase transitions and ionic motions in solid trimethylethylammonium iodide studied by ^1H and ^{127}I NMR, electrical conductivity, X-ray diffraction and thermal analysis. *Ber. Bunsenges. Phys. Chem.* **100**, 433–439 (1996).
- Tanabe, T., Nakamura, D. & Ikeda, R. Novel ionic plastic crystal phase of $[(\text{CH}_3)_4\text{N}]\text{SCN}$ obtainable above 455K studied by proton magnetic resonance, electrical conductivity and thermal measurements. *J. Chem. Soc. Faraday Trans.* **87**, 987–990 (1991).
- Shimizu, T., Tanaka, S., Onoda-Yamamoto, N., Ishimaru, S. & Ikeda, R. New rotator phase revealed in di-n-alkylammonium bromides studied by solid state NMR, powder XRD, electrical conductivity and thermal measurements. *J. Chem. Soc. Faraday Trans.* **93**, 321–326 (1997).
- Cooper, E. I. & Angell, C. A. Ambient temperature plastic crystal fast ion conductors (PLICFICS). *Solid State Ionics* **18/19**, 570–576 (1986).
- Aronson, R. et al. Fast ion conductors with rotating sulphate ions. *J. Phys. Colloq. C6* **41**, 35–37 (1980).
- Aronson, R., Knappe, H. E. G. & Torell, L. M. Brillouin spectra of the solid electrolyte Li_2SO_4 . *J. Chem. Phys.* **77**, 677–680 (1982).
- Borgesson, L. & Torell, L. M. Reorientational motion in superionic sulphates: A Raman linewidth study. *Phys. Rev. B* **32**, 2471–2477 (1985).
- MacFarlane, D. R., Sun, J., Forsyth, M., Meakin, P. & Amini, N. Pyrrolidinium imides: a new family of molten salts and conductive plastic crystal phases. *J. Phys. Chem.* **103**, 4164–4170 (1999).
- Sun, J., Forsyth, M. & MacFarlane, D. R. Room temperature molten salts based on the quaternary ammonium ion. *J. Phys. Chem.* **102**, 8858–8864 (1998).
- Sylla, S., Sanchez, J.-Y. & Armand, M. Electrochemical study of linear and crosslinked POE-based polymer electrolytes. *Electrochim. Acta* **37**, 1699–1701 (1992).

Correspondence and requests for materials should be addressed to D.R.M. (e-mail: D.MacFarlane@sci.monash.edu.au).

Patterning liquid flow on the microscopic scale

Dawn E. Kataoka & Sandra M. Troian

Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544-5263, USA

Microscopic fluidic devices, ranging from surgical endoscopes¹ and microelectromechanical systems² to the commercial 'lab-on-a-chip' (ref. 29), allow chemical analysis and synthesis on scales unimaginable a decade ago. These devices transport minuscule quantities of liquid along networked channels. Several techniques have been developed to control small-scale flow, including micromechanical³ and electrohydrodynamic⁴ pumping, electro-osmotic flow⁵, electrowetting^{6,7} and thermocapillary pumping^{8–10}. Most of these schemes require micro-machining of interior channels and kilovolt sources to drive electrokinetic flow. Recent work^{8–10} has suggested the use of temperature instead of electric fields to derive droplet movement. Here we demonstrate a simple, alternative technique utilizing temperature gradients to direct microscopic flow on a selectively patterned surface (consisting of alternating stripes of bare and coated SiO_2). The liquid is manipulated by simultaneously applying a shear stress at the air–liquid interface and a variable surface energy pattern at the liquid–solid interface. To further this technology, we provide a theoretical estimate of the smallest feature size attainable with this technique.

A thin, wetting film in contact with a solid substrate bearing a temperature gradient will flow from warmer to cooler regions because of the development of a thermocapillary shear stress at the air–liquid interface¹¹. For sufficiently thin films in which the thermal conductive resistance is much smaller than the convective resistance (that is, at small Biot number), the applied thermal gradient at the solid surface is directly transferred to the air–liquid interface. Because most liquids maintain a negative and constant value of $d\gamma/dT$, where γ is the surface tension and T is the temperature, the application of a constant thermal gradient in

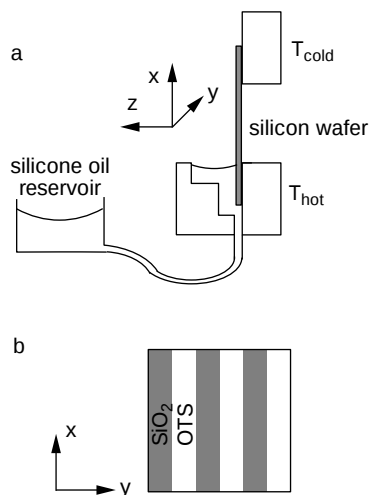


Figure 1 Schematic diagrams of the experimental apparatus (**a**) and the orientation of the patterned Si wafer (**b**). The wafer was patterned by spin-coating photoresist onto an OTS-coated wafer for 40 s at 4,000 r.p.m., and baking for 20 min at 90 °C. Using a mask aligner, the wafer was then exposed to UV light through a mask consisting of vertical stripes. The wafer was placed in a developer which stripped away the exposed photoresist, baked at 90 °C for 20 min, then subjected to an oxygen plasma for 20 s; the plasma removed any remaining OTS which lay beneath the developed photoresist. The wafer surface now consisted of alternating stripes of bare, oxidized SiO_2 and photoresist-covered OTS. A final acetone rinse removed the undeveloped photoresist, leaving the OTS monolayer intact. Contact angle measurements with water on surfaces patterned at a larger scale indicated that the uncoated silicon portions were completely hydrophilic, while the OTS-coated portions produced an angle $>110^\circ$, as required for uniform OTS monolayers^{26,27}.

the streaming ($\hat{x}$) direction produces a fixed shear stress at the air–liquid surface given by $\tau = d\gamma/dx = (d\gamma/dT)(dT/dx)$. Films that are driven by a thermocapillary shear stress can even climb a vertical substrate, provided that the upward flux due to thermal stresses is larger than the downward flux due to gravitational drainage (that is, $h_0 \ll 3\tau/2\rho g$, where h_0 is the film thickness, ρ is the fluid density and g is the gravitational acceleration). In systems with large surface-to-volume ratio, as in microscale flows, even small thermal gradients can produce significant liquid movement.

Thermocapillary spreading on homogeneous surfaces has been shown to undergo a hydrodynamic instability for sufficiently thin films or large surface stress^{12–16}. In this limit, the spreading front develops a thickened capillary ridge; for a certain range of transverse wavenumbers, this ridge becomes unstable to the formation of numerous, parallel rivulets whose spacing is governed by a characteristic wavelength, $\lambda_c \propto h_0(3Ca)^{-1/3}$. The capillary number, $Ca = \eta V/\gamma$, represents the balance between the viscous, capillary and thermocapillary force where η is the liquid viscosity and V is the steady-state spreading velocity. For films driven by this thermocapillary effect¹⁴, $Ca = \tau h_0/2\gamma$ and $\lambda_c = 18(2\gamma h_0^2/3\tau)^{1/3}$. Energy decomposition techniques have identified the source of instability^{14–16}: disturbances which cause a local protrusion in the spreading front create thickened portions which spread faster than adjacent thinner regions, thereby destabilizing an initially uniform front. This same mechanism, first described by Spaid and Homsy¹⁷, is responsible for the fingering instability observed in gravitational^{18–22} and centrifugal^{17,23,24} flows.

From a technological point of view, the possibility of using this spontaneous pattern formation to channel microvolumes of liquid along a substrate is questionable. For spreading on homogeneous surfaces, the rivulets almost never appear in the same position along the spreading front. In addition, surface defects or contaminant particles tend to produce flow irregularities which create as much as

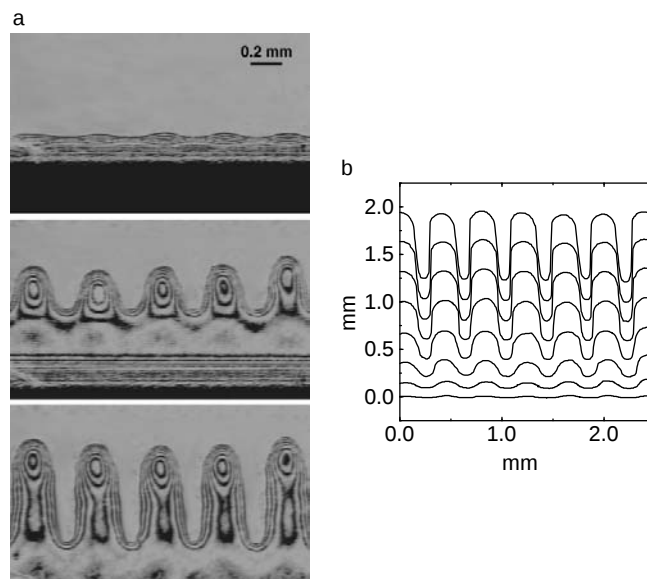


Figure 2 Time series of a silicone-oil film spreading on a patterned silicon wafer. The pattern consists of alternating 200-μm-wide stripes of OTS and bare oxidized SiO_2 , and the spreading was monitored by interferometry. **a**, Three interferograms, with a five minute lapse between each. Adjacent fringes correspond to a change in film thickness of 0.2255 μm. **b**, A series of digitized curves, with a two minute lapse between each, representing the liquid front. The applied shear stress for **a** and **b** is $\tau = 0.8 \text{ dyn cm}^{-2}$.

a 30% variation in the natural wavelength, λ_c (refs 21, 22, 25). There exists, however, a special feature of thermally driven films which can be manipulated to regularize the flow along selected, prescribed lanes. Even small thermal gradients produce liquid films less than a micrometre in thickness^{12,13,25}. These films therefore maintain an excessively large surface to volume ratio. The combination of an applied shear stress at the air–liquid surface and a patterned surface energy profile at the liquid–solid interface can be used to direct microscale flow along prescribed lanes. The flows produced in this manner are stable and reproducible, as the surface patterns control accessible or non-accessible regions. Because of the strong influence of the solid surface, the flow also rapidly self-corrects in the presence of defects or surface contaminants, always keeping perfect registry with the surface pattern. From a practical and fundamental point of view, it is therefore interesting to explore the consequences of patterning the surface with a wavelength different from the natural wavelength which would be selected by the liquid were it to spread on a homogeneous surface. Our results show that the instability wavelength can be easily superseded by the substrate pattern, provided that the width of the liquid channels exceeds a lower critical value.

We have examined the thermocapillary spreading of a silicone oil films (PDMS, both with viscosity of 18 mPa s and surface tension of 21 dyn cm^{-1} at 25 °C, Aldrich) on vertically oriented, patterned silicon wafers. The wafers were held vertical by suction against two brass blocks maintained at different temperatures by two circulating baths (see Fig. 1a). The thermal gradient was recorded after each experimental run by six equally spaced, fine wire thermocouples (0.005 inch diameter, copper/constantin, Omega). For the temperature range we studied, silicone oil maintains a value of $d\gamma/dT = -0.05 \text{ dyn cm}^{-1} \text{ } ^\circ\text{C}^{-1}$. When placed in contact with the heated portion of the substrate, the silicone oil spontaneously climbs along the wafer surface. The brass block at the higher temperature was fed by a small Teflon tube connected to a reservoir of silicone oil mounted on a vertical stage (Fig. 1a). By varying the height of the stage, the meniscus level on the wafer could be adjusted precisely. In these experiments, the shear stress, τ , was in the range 0.73–0.80 ($\pm 10\%$) dyn cm^{-2} .

The surface pattern chosen for study was a sequence of alternating vertical stripes whose wavelength ranged from 50 to 500 μm . The vertical stripes consisted either of bare, naturally oxidized silicon (SiO_2), or oxidized silicon coated with a covalently bonded monolayer of octadecyltrichlorosilane (OTS). The orientation of the patterned wafer is shown in Fig. 1b. The wafers were cleaned, and then immersed in an OTS solution under a nitrogen atmosphere, according to published procedures²⁶. OTS monolayers prepared by this method have been shown to produce a stable and complete monolayer of 26 Å thickness²⁷. The coated wafers were then patterned by photolithography as described in Fig. 1 legend. This same procedure works equally well on glass substrates. We also tried patterning the OTS monolayer using deep ultraviolet (193 nm) illumination through a fused silica mask²⁸; this one-step procedure also produced excellent results.

The topography of the spreading film was visualized by laser interferometry. Light from a He–Ne laser (wavelength 6,328 Å) impinged on the substrate at normal incidence, producing the interferograms shown in Fig. 2a. The interfringe spacing corresponds to a change in film thickness of 0.2555 μm . Typical film thicknesses were less than 1 μm for these shear stresses. The images were used to monitor the speed of the advancing front, and to reconstruct the entire surface shape. The images were recorded with a high-resolution CCD camera (MTI VE1000, Dage) fitted with a Navitar zoom lens and stored on SVHS tapes for image analysis. Further details can be found in ref. 25.

We first examined the spreading behaviour of silicone oil on homogeneous surfaces consisting either of bare, oxidized SiO_2 wafers, or wafers fully coated with OTS. Contact angle and work of adhesion measurements for silicone oil on either substrate revealed a very small and similar contact angle. (According to Young's equation, however, this does not imply that the liquid–solid surface energy is the same.) The spreading on either homogeneous surface proceeded by the instability mechanism described earlier, with a well defined natural wavelength set by λ_c . The spreading velocities were also similar.

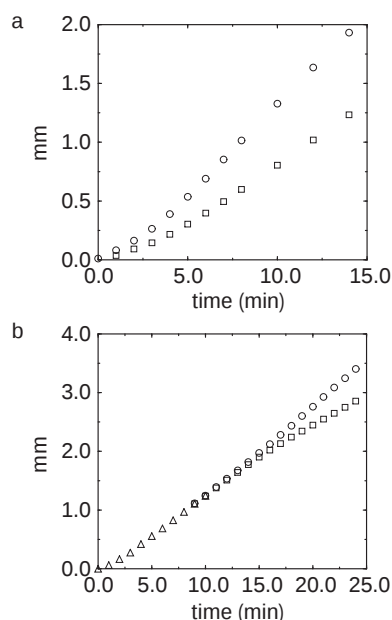


Figure 3 Comparison between the location of the liquid front for spreading on a patterned and an homogeneous surface. Circles and squares represent liquid channel tips and troughs, respectively. **a**, Silicone oil spreading on a wafer patterned with alternating 200- μm -wide stripes of OTS and bare oxidized SiO_2 . **b**, Silicone oil spreading on a homogeneous, bare, oxidized SiO_2 wafer. The applied shear stress in **a** and **b** is $\tau = 0.8 \text{ dyn cm}^{-2}$.

Experiments with the patterned wafers revealed a very different flow behaviour. Figure 2a shows a time sequence of the liquid film spreading on a wafer bearing alternating 200- μm -wide stripes of OTS and uncoated SiO_2 . Whereas (unstable) rivulet formation on homogeneous surfaces always requires an incubation period during which the front spreads as a straight line before fingering, the flow on patterned surfaces is immediately channelled into well defined lanes whose spacing is imposed by the surface pattern (Fig. 2b). In Fig. 3, we have plotted the advance of the spreading front on a patterned and homogeneous substrate (bare SiO_2) with all other variables remaining constant. The spreading on the patterned wafers (Fig. 3a) immediately creates alternating tips (circles) and troughs (squares). The homogeneous surface instead allows uniform spreading for over 15 min before the front becomes unstable (Fig. 3b), spontaneously splitting into tips (circles) and troughs (squares). For a shear stress $\tau = 0.8 \text{ dyn cm}^{-2}$, the average speed of the tips and troughs on the patterned wafer was 2.60 $\mu\text{m s}^{-1}$ and 1.73 $\mu\text{m s}^{-1}$ and on the bare wafer 2.56 $\mu\text{m s}^{-1}$ and 1.94 $\mu\text{m s}^{-1}$, respectively. The tips advanced at similar speed on either surface, while the troughs appeared to advance more slowly on the OTS segments.

In Table 1 we summarize the results for spreading on homogeneous and patterned surfaces with variable width strips. In all cases, the silicone oil preferentially spread on the uncoated SiO_2 portions. This is not surprising, as a clean silicon wafer covered by natural oxide layer has a high surface energy and therefore a high affinity for water and organic contaminants. Organic monolayers terminating in methyl and vinyl groups have low surface energy, and resist organic contamination²⁷. For the homogeneous surfaces, the periodicity agrees well with the predicted value for λ_c . On the patterned surfaces, the periodicity of the spreading front, λ_{exp} , agrees well with the periodicity imposed by the solid substrate, λ_{solid} . The excellent registry with the imposed surface pattern persisted for all flows except the last entry. In this last example, the liquid was unable to follow the narrow surface lanes and reverted back to the natural wavelength that would be obtained on a homogeneous surface.

A simple scaling argument predicts the approximate lane width for which the capillary pressure at the tip of a liquid channel is strong enough to oppose the upward thermally driven flow for patterned spreading. In this limit, the upward volumetric flux, $\tau h_0^3/2\eta$, is opposed by the downward capillary flux, $h_0^3(dp/dx)/3\eta$, where p is the Laplace pressure. For a completely wetting liquid, $(\nabla h_0)^2 \ll 1$, $dp/dx \sim O(\gamma h_0/R^3)$, where R denotes the radius of curvature in the streaming direction. This balance produces a lower critical value of $R \sim (\gamma h_0^2/\tau)^{1/3}$. Strong curvature in the transverse (y) direction, which induces lateral flow, can also force liquid to spread over the OTS stripes. However, without an estimate for the pinning force exerted by the junction between the OTS and uncoated stripes, the corresponding force balance cannot be expressed. Nonetheless, as seen in Fig. 2, the radius of curvature

Table 1 Liquid lane spacing for silicone oil spreading on homogeneous and patterned surfaces

τ (dyn cm^{-2})	W_{OTS} (μm)	W_{SiO_2} (μm)	λ_{solid} (μm)	λ_{exp} (μm)
0.79	—	All	—	574
0.79	All	—	—	545
0.77	250	250	500	489
0.80	200	200	400	399
0.76	150	150	300	288
0.73	250	150	400	393
0.73	150	250	400	394
0.74	50	50	100	545

The first two rows show data for fully uncoated and coated surfaces, respectively. τ is the applied surface stress, W_{OTS} and W_{SiO_2} denote the widths of each type of lane, $\lambda_{\text{solid}} = W_{\text{OTS}} + W_{\text{SiO}_2}$ denotes the wavelength of the repeating surface pattern, and λ_{exp} represents the average measured distance between adjacent finger tips. The distance, λ_{exp} , was obtained from an edge-detection algorithm which averaged over 10–20 liquid fingers.

in both the streaming and transverse direction is fairly similar, so that this estimate for R provides a reliable lower cut-off. This critical value can be made even smaller by decreasing γ (using different liquids or introducing surfactants), and by decreasing h_0 or increasing τ . Typical values for the film thickness ($1\ \mu\text{m}$), surface tension ($20\ \text{dyn cm}^{-1}$), and shear stress ($1\ \text{dyn cm}^{-2}$) yields a characteristic finger or lane width, $2R \sim O(100\ \mu\text{m})$, in good agreement with the flow behaviour observed in the last entry of Table 1.

To develop further this technology for microfluidic applications, the troughs of the liquid front can be pinned in place, allowing liquid to flow only along specified surface lanes. This can be accomplished by either decreasing the surface energy of adjacent lanes (so that the liquid becomes partially or even non-wetting along those stripes) or by micromachining flow obstacles at the liquid entrance point. We are experimenting with different monolayer treatments for this purpose. We have also patterned more complicated surface networks to force liquid to spread in curved lanes and around corners by the application of thermal gradients in and across the streaming direction. In this way, we can achieve precise flow control at the microscopic scale by simultaneously manipulating the boundary conditions at the air-liquid and liquid-solid interfaces. These concepts should be generalizable to other driving forces which are proportional to the surface area of the fluid, not to its volume. $\square$

Received 23 March; accepted 4 October 1999.

1. Menz, W. & Guber, A. Microstructure technologies and their potential in medical applications. *Min. Invasive Neurosurgery* **37**, 21–27 (1994).
2. Ho, C. H. & Tai, Y. C. Micro-electro-mechanical-systems (MEMS) and fluid flows. *Annu. Rev. Fluid Mech.* **30**, 579–612 (1998).
3. van Lintel, H. T. G. A piezoelectric micropump based on micromachining of silicon. *Sensors Actuators* **15**, 153–167 (1988).
4. Bart, S. F., Tarrow, L. S., Mehregany, M. & Lang, J. H. Microfabricated electrohydrodynamic pumps. *Sensors Actuators A* **21–23**, 193–197 (1990).
5. Manz, A. *et al.* Electroosmotic pumping and electrophoretic separations for miniaturized chemical analysis systems. *J. Micromech. Microeng.* **4**, 257–265 (1995).
6. Beni, G. & Tenen, M. A. Dynamics of electrowetting displays. *J. Appl. Phys.* **52**, 6011–6015 (1981).
7. Gallardo, B. S. *et al.* Electrochemical principles for active control of liquids on submillimeter scales. *Science* **283**, 57–59 (1999).
8. Handique, K., Gogoi, B. P., Burke, D. T., Mastrangelo, C. H. & Burns, M. A. Microfluidic flow control using selective hydrophobic patterning. *Proc. SPIE* **3224**, 185–195 (1997).
9. Handique, K., Burke, D. T., Gogoi, B. P., Mastrangelo, C. H. & Burns, M. A. in *Technical Digest: Solid State Sensor and Actuator Workshop* 346–349 (Transducer Research Foundation, Cleveland, Ohio, 1998).
10. Sammarco, T. S. & Burns, M. A. Thermocapillary pumping of discrete drops in microfabricated analysis devices. *Am. Inst. Chem. Eng. J.* **45**, 350–366 (1999).
11. Ludviksson, V. & Lightfoot, E. N. The dynamics of thin liquid films in the presence of surface-tension gradients. *Am. Inst. Chem. Eng. J.* **17**, 1166–1173 (1971).
12. Cazabat, A. M., Heslot, F., Troian, S. M. & Carles, P. Fingering instability of thin spreading films driven by temperature gradients. *Nature* **346**, 824–826 (1990).
13. Brzoska, J. B., Brochard-Wyart, F. & Rondelez, F. Exponential growth of fingering instabilities of spreading films under horizontal thermal gradients. *Europhys. Lett.* **19**, 97–102 (1992).
14. Kataoka, D. E. & Troian, S. M. A theoretical study of instabilities at the advancing front of thermally driven coating films. *J. Colloid Interface Sci.* **192**, 350–362 (1997).
15. Kataoka, D. E. & Troian, S. M. Stabilizing the advancing front of thermally driven coating films. *J. Colloid Interface Sci.* **203**, 335–344 (1998).
16. Kataoka, D. E. & Troian, S. M. The transient dynamics and structure of optimal excitations in thermally driven films. *Phys. Fluids*. (submitted).
17. Spaid, M. A. & Homay, G. M. Stability of Newtonian and viscoelastic dynamic contact lines. *Phys. Fluids* **8**, 460–478 (1996).
18. Huppert, H. E. Flow and instability of a viscous current down a slope. *Nature* **300**, 427–429 (1982).
19. Silvi, N. & Dussan, E. B. On the rewetting of an inclined solid surface by a liquid. *Phys. Fluids* **28**, 5–7 (1985).
20. Troian, S. M., Herbolzheimer, E., Safran, S. A. & Joanny, J. F. Fingering instabilities of driven spreading films. *Europhys. Lett.* **10**, 25–30 (1989).
21. Jerrett, J. M. & deBruyn, J. R. Finger instability of a gravitationally driven contact line. *Phys. Fluids A* **4**, 234–242 (1992).
22. deBruyn, J. R. Growth of fingers at a driven three phase contact line. *Phys. Rev. A* **46**, R4500–R4503 (1992).
23. Melo, F., Joanny, J. F. & Fauve, S. Fingering instability of spinning drops. *Phys. Rev. Lett.* **63**, 1958–1961 (1989).
24. Frayssé, N. & Homay, G. M. An experimental study of rivulet instabilities in centrifugal spin coating of viscous Newtonian and non-Newtonian fluids. *Phys. Fluids* **6**, 1491–1504 (1994).
25. Kataoka, D. E. *The Spreading Behavior of Thermally Driven Liquid Films*. Thesis, Princeton Univ. (1999).
26. Brzoska, J. B., Ben Azouz, I. & Rondelez, F. Silanization of solid substrates: a step toward reproducibility. *Langmuir* **10**, 4367–4373 (1994).
27. Wasserman, S. R. *et al.* The structure of self-assembled monolayers of alkylsiloxanes on silicon: a

comparison of results from ellipsometry and low-angle x-ray reflectivity. *J. Am. Chem. Soc.* **111**, 5852–5861 (1989).

28. Dulcey, C. L. *et al.* Deep UV photochemistry of chemisorbed monolayers: patterned coplanar molecular assemblies. *Science* **252**, 551–554 (1991).
29. Freemantle, M. Downsizing chemistry. *Chem. Eng. News* **77**(8), 27–36 (1999).

Acknowledgements

We thank G. D. Barriac for assisting with the photolithographic patterning of the silicon wafers. This work was supported by the National Science Foundation through a graduate fellowship (D.E.K.) and a CAREER award (S.M.T.).

Correspondence and requests for materials should be addressed to S.M.T. (e-mail: stroian@princeton.edu).

Stabilizing a solid–solid interface with a molecular-scale adhesive

R. J. Davey, L. Williams-Seton, H. F. Lieberman & N. Blagden

Colloids, Crystals and Interfaces Group, Department of Chemical Engineering, UMIST, Manchester M60 1QD, UK

The industrial importance of molecular materials chemistry has promoted great interest in areas such as self-assembled surface coatings¹, multi-layer formation on solid substrates², crystallization from solutions³, crystal morphology⁴ and structure prediction⁵, solving structures from powders⁶ and control of polymorphism⁷. Improvements in our understanding of the role of intermolecular interactions in driving molecular self-assembly and interfacial processes have led to technological advances—both in controlling the assembly of molecules at the nanometre scale, and in manipulating processes and products in which crystal nucleation and growth are key elements⁸. But there has been relatively little work on molecular-scale engineering at solid–solid interfaces, despite their importance in polymeric

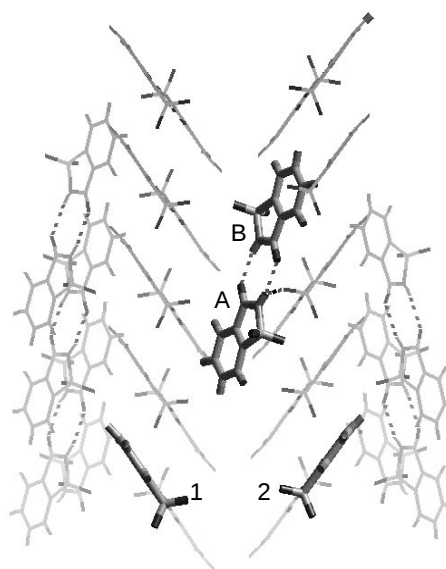


Figure 1 The $(10\bar{2})$ twin interface in saccharin viewed normal to $(10\bar{2})$. Molecules A and B, lying in the twin plane, illustrate how the structure is created from hydrogen-bonded dimers. Molecules 1 and 2, appearing side on, are related across the twin interface by mirror symmetry.

composites for structured and electronic applications, in adhesives and in formulated pharmaceutical and agrochemical products. Here we report the use of molecules as tailored adhesives—a molecular ‘glue’ is selected to bond across an interfacial region and hence stabilize a solid–solid interface. We consider a simple interface occurring in a twinned crystal of saccharin; additive molecules with predictable dimensions and hydrogen-bonding functionality can span the interface. The stabilization is reflected in an enhanced frequency of twin-crystal formation.

In a twinned crystal, the twin plane is a solid–solid interface comprising identical molecules arranged in the same crystal structure, and with the twin portions related by a simple mirror or rotational symmetry. One significant problem even in such a well defined case lies in knowing the precise molecular arrangement either side of the twin plane: while conventional crystallography can be used to define the twin plane and the twin operation, it cannot yield atomic coordinates of the few layers of molecules adjacent to the twin interface. Without known molecular positions, molecular-scale manipulation of such an interface is impossible. In the work we report here, we were able to use the description of the saccharin twin interface determined in our previous studies, which showed that the twin plane is (10 $\bar{2}$) and that the two halves of the macroscopic crystal are related by mirror symmetry⁹. The precise arrangement of molecules about this plane was determined through energy calculation, by finding the juxtaposition of the twinned portions for which the energy of the twinned crystal was minimized⁹. This was found to coincide with a partial (pseudo) glide plane, relating the two halves of the twin and the formation of a new three-centre hydrogen bond across the interface, an arrangement we have shown to be consistent with the solvent dependency of twin formation¹⁰. Here we apply this model of the interface to the problem of selecting auxiliary molecules to bridge the twin interface and act as a molecular ‘glue’.

Visualized using CERIUSt¹¹ software, the interface is shown in Fig. 1. This figure illustrates firstly how in the bulk structure, saccharin molecules (see Fig. 1 legend) pack in a dimer motif, utilizing intermolecular amide hydrogen bonds, and secondly how these dimers pack at the twin interface (molecules A and B highlighted) interacting via a new three-centre hydrogen bond. It is also clear from this model that the molecules 1 and 2, related by the glide plane of the twin and lying on either side of the interface, both expose sulphoxide oxygens, separated from each other by a distance of 0.58 nm, to the internal interfacial region. Because in the existing crystal these oxygens are not utilized for intermolecular interactions, we sought to use them as points of intermolecular contact with potential adhesive molecules. In selecting adhesives, therefore, we chose molecules which could bridge the twin interface by hydrogen-bonding between these oxygen atoms, hence both replacing the

central dimer pair (molecules A and B) and further stabilizing the interface via the new intermolecular contacts. This concept is shown schematically in Fig. 2, which highlights the double requirement of matching both dimension and intermolecular interaction across the interfacial region.

On the above basis, we selected three molecules to test; ethylene glycol, indigo and hydroquinone. The separation of the hydroxyl hydrogens in the *trans* configuration of ethylene glycol is about 0.47 nm, and in indigo the separation of thiazole nitrogens (those in the two adjacent five-membered rings, Fig. 3) is similarly 0.47 nm. Both these molecules are thus capable of sitting across the twin plane and stabilizing the twin interface, as seen in Fig. 3. A priori we expected the conformational rigidity of indigo to enhance its efficacy relative to ethylene glycol. In addition, we anticipated that as a further test of our model, indigo would be expected to dye the interface blue, and hence provide a visual indication of its location in the final twinned crystals. As a third molecule we chose 1,4 dihydroxybenzene (hydroquinone), a molecule which is essentially rigid, with the hydroxyl hydrogens separated by 0.62 nm. On the basis of our model, this molecule is too large to adopt a position across the interface without creating significant, destabilizing steric clashes. We expected it therefore to destabilize twin formation.

In order to test these hypotheses we performed simple cooling crystallization experiments⁹, in which the selected molecules were dissolved in ethanolic solutions of saccharin (supersaturation 0.4) at a concentration of 2.7×10^{-4} M. The extent to which the measured frequency of twinning was changed relative to the pure solution was then taken as a measure of the efficacy of the chosen adhesive molecules in stabilizing the solid–solid interface. We judged that this would provide a rigorous test of our strategy and selection protocol for the following reasons. First, the occurrence of twinned crystals in pure ethanol solutions was low, and independent of supersaturation; second, because twinning always occurred at the point of nucleation^{9,10}, yielding hemihedral twins. These are important mechanistic points as they imply that the kinetic processes associated with crystal growth play no role in twin formation. This means that our conclusions will not be influenced by the potential effects that these additives may have on growth-related processes such as habit modification or crystal perfection.

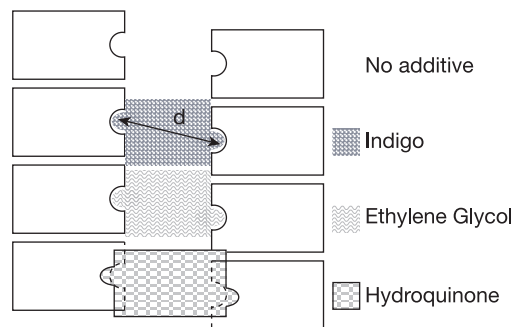


Figure 2 Schematic diagram showing the mechanism of the action of the additives. The distance d between the hydrogen-bonding centres must be suitable for the match to occur. For indigo and ethylene glycol d is 0.47 nm; for hydroquinone, d is 0.62 nm which is too large to fit across the interface.

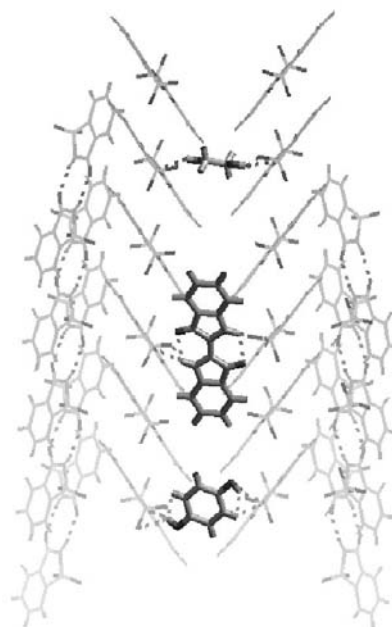


Figure 3 The (10 $\bar{2}$) interface showing positions of additives. Top to bottom: ethylene glycol, indigo and hydroquinone are shown docked across the interface.

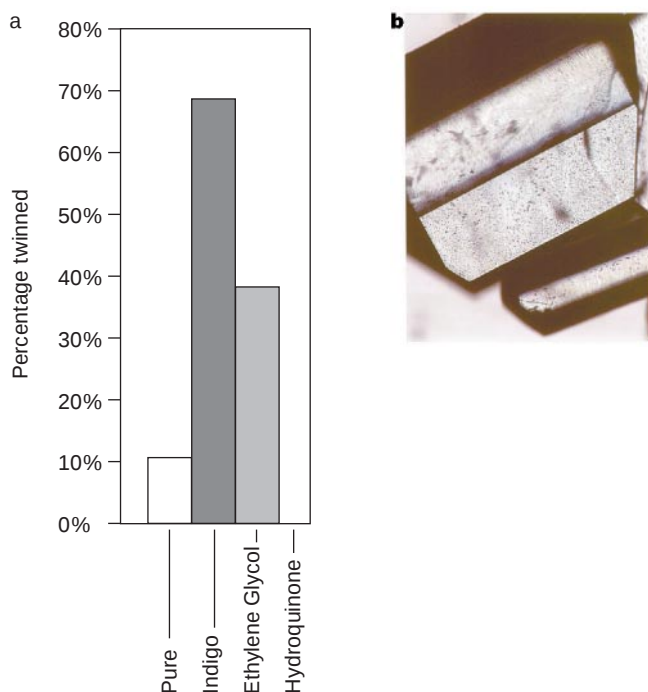


Figure 4 Production of twinned saccharin crystals. **a**, A comparison of the frequency of twinning in pure and doped solutions for crystals grown at 25 °C and a supersaturation of 0.4 in unstirred vessels. **b**, Micrograph of a twinned saccharin crystal grown from an indigo-doped saccharin solution, showing blue colouration at the interface.

After crystallization, crystals were removed and the percentage of twinned crystals determined by manual counting. The number of crystals counted varied between experiments, but was always between 100 and 200. The histogram in Fig. 4a summarizes these results: in pure solutions, 10% of crystals grow as twins; in the presence of ethylene glycol, 40% grow as twins; while in solutions containing indigo, almost 70% of the crystals are twinned. These data are in line with our expectations. In addition, we show in Fig. 4b an optical micrograph of a detail of the twin interface for a crystal grown in the presence of indigo. The blue colouration verifies that this molecule is indeed concentrated at the interface. In complete contrast, the addition of hydroquinone was found to yield no twins at all, again confirming our expectations.

The success of this study shows that our understanding of intermolecular interactions and interfacial processes can be applied successfully to the solid–solid interface. It suggests a general strategy for designing adhesive molecules to span and stabilize such interfaces, utilizing a combination of molecular dimensions and specific intermolecular interactions. We note that we were only able to demonstrate this success because we had previously developed a molecular model of the interfacial region—without this, such engineering would be relegated to randomized trials and serendipitous successes. Fortunately, such information is rapidly becoming routinely available, not only through the increased power and sophistication of molecular modelling, but also using state-of-the-art structural techniques such as scanning probe microscopy, microscopic spectroscopy and glancing-angle X-ray diffraction¹². □

Received 7 June; accepted 19 October 1999.

1. Maoz, R., Matlis, S., Dimasi, E., Ocko, B. M. & Sagiv, J. Self-replicating amphiphilic monolayers. *Nature* **384**, 150–153 (1996).
2. Hatzor, A. *et al.* Coordination-controlled self-assembled multilayers on gold. *J. Am. Chem. Soc.* **120**, 13469–13477 (1998).
3. Weissbuch, I., Popvitz-Biro, R., Leiserowitz, L. & Lahav, M. in *The Lock and Key Principle* (ed. Behr, J. P.) 173–246 (Wiley & Sons, Chichester, 1994).
4. Docherty, R. & Roberts, K. J. Modelling the morphology of molecular crystals. *J. Cryst. Growth* **88**, 159–168 (1988).

5. Gavazotti, A. & Fillipini, G. Polymorphic forms of organic crystals at room conditions: thermodynamic and structural implications. *J. Am. Chem. Soc.* **117**, 12299–12305 (1995).
6. Docherty, R. & Jones, W. in *Organic Molecular Solids—Properties and Applications* (ed. Jones, W.) Ch. 5 (CRC Press, London, 1997).
7. Davey, R. J., Blagden, N., Potts, G. D. & Docherty, R. Polymorphism in molecular crystals: stabilization of a metastable form by conformational mimicry. *J. Am. Chem. Soc.* **119**, 1767–1772 (1997).
8. Davey, R. J., Maginn, S. J. & Polywka, L. in *Advances in Industrial Crystallisation* 150–165 (Butterworth-Heinemann, Oxford, 1991).
9. Lieberman, H. F., Williams, L., Davey, R. J. & Pritchard, R. G. Molecular configuration at the solid–solid interface: twinning in saccharin crystals. *J. Am. Chem. Soc.* **120**, 686–691 (1998).
10. Williams-Seton, L., Davey, R. J. & Liebermann, H. F. Solution chemistry and twinning in saccharin crystals: a combined probe for the structure and functionality of the crystal–fluid interface. *J. Am. Chem. Soc.* **121**, 4563–4567 (1999).
11. Cerius² version 2.0 (Molecular Simulations Inc., Cambridge, UK, 1998).
12. Gidalevitz, D. *et al.* *Angew. Chem. Int. Edn Engl.* **36**, 955–959 (1997).

Acknowledgements

L.W.-S. and H.F.L. acknowledge support from ROPA awards.

Larval retention and recruitment in an island population of a coral-reef fish

Stephen E. Swearer*, Jennifer E. Caselle*, David W. Lea† & Robert R. Warner*

*Department of Ecology, Evolution and Marine Biology and Marine Science Institute, †Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA

For close to a century, recruitment of larvae to a local population has been widely accepted as a primary determinant of marine population dynamics^{1,2}. However, progress in elucidating the causes of recruitment variability has been greatly impeded by our ignorance of the sources of recruits. Although it is often assumed that recruitment is independent of local reproduction^{3–6}, there is increasing circumstantial evidence that physical^{7,8} and behavioural^{9,10} mechanisms could facilitate larval retention near source populations. To develop a direct method for reconstructing the dispersal history of recruiting larvae, we put forward the hypothesis that differences in nutrient and trace-element concentrations between coastal and open oceans could result in quantifiable differences in growth rate and elemental composition between larvae developing in coastal waters (locally retained) and larvae developing in open ocean waters (produced in distant locations). Using this method, we show that recruitment to an island population of a widely distributed coral-reef fish may often result from local retention on leeward reefs. This result has implications for fisheries management and marine reserve design, because rates of dispersal between marine populations—and thus recruitment to exploited populations—could be much lower than currently assumed.

Our approach is based on the following two predictions. (1) Coastal environments are the most productive regions of the world's oceans¹¹, where local nutrient inputs can enhance food resources for developing fish larvae^{12,13}. Thus, larvae retained in coastal water masses should grow faster, and settle at larger sizes, than larvae dispersing through nutrient-depleted oceanic waters. (2) Along coastal margins, concentrations of many particle-reactive trace elements are significantly higher than in the open ocean¹⁴. Thus, larvae developing in coastal environments should have elevated trace-element concentrations in their tissues compared with oceanic larvae if these elements are absorbed from sea water and permanently incorporated into the larvae. We used otoliths—crystalline structures in the inner ear of teleost fishes—as recorders of growth history and trace-element environmental signatures because they are formed at birth, new layers of calcium carbonate

are deposited daily as discernible growth bands¹⁵, and elements from the surrounding sea water can substitute for calcium and become permanently incorporated into the otolith^{16,17}. In addition, otolith trace-element signatures have recently been used to identify stocks¹⁷ and nursery habitats^{18,19} for several temperate fishes.

The focal species for this study was the bluehead wrasse, *Thalassoma bifasciatum*, an abundant Caribbean coral-reef fish with a long planktonic larval duration (~45 days). Past research on St Croix, US Virgin Islands (northeastern Caribbean Sea, 17° 45' N, 64° 35' W) has detected consistent recruitment patterns²⁰. On the windward shore of St Croix, recruitment was highest on the eastern (up-current) end and decreased towards the western end of the island. Along the leeward shore, recruitment was lowest at the eastern end and increased to highest recruit densities at the westward (down-current) end of the island. We put forward the hypothesis that physical oceanographic processes were responsible for the patterns of recruitment along the two shores. Windward reefs may receive recruits primarily from larvae dispersing from nearby up-current source populations (possibly Saba, St Martin, or other islands of the lesser Antilles, from where passively drifting larvae could reach St Croix in as little as 7–14 days; ref. 21) whereas leeward reefs may receive recruits retained within coastal waters in the lee of the island. Regions of weak converging currents as well as wake eddies have been detected along the leeward coast of St Croix²², and may also facilitate larval retention near other islands^{23,24}. Thus the importance of larval dispersal versus local retention to patterns of recruitment may vary across windward and leeward reefs.

In order to determine the sources of recruits to St Croix, we compared larval growth histories and otolith trace-element composition (see Methods) of newly settled individuals among three locations: an up-current site on the windward shore (Jack's Bay) and two down-current sites on the leeward shore (Northstar and Butler Bay) during the summer and autumn of 1992. For this analysis, we selected individuals that recruited during five periods: three summer periods (June–August) when average recruitment was more than five times greater at the leeward sites relative to the windward site, and two autumn periods (September and October) when average recruitment levels were lower and similar among the leeward and windward shores (1.1 and 0.9 fish per m², respectively)²⁰. We found significant differences in both growth history and otolith trace-element signatures among larval populations of bluehead wrasse recruiting to each site as well as among recruitment periods (Table 1), indicating that recruiting larvae experienced different environmental histories both within recruitment periods as well as among recruitment periods.

Differences among recruitment periods resulted from the clustering of both larval growth history and otolith trace-element signatures along a retention–dispersal gradient as indicated by canonical discriminant analysis (Fig. 1; see Methods). In support of our

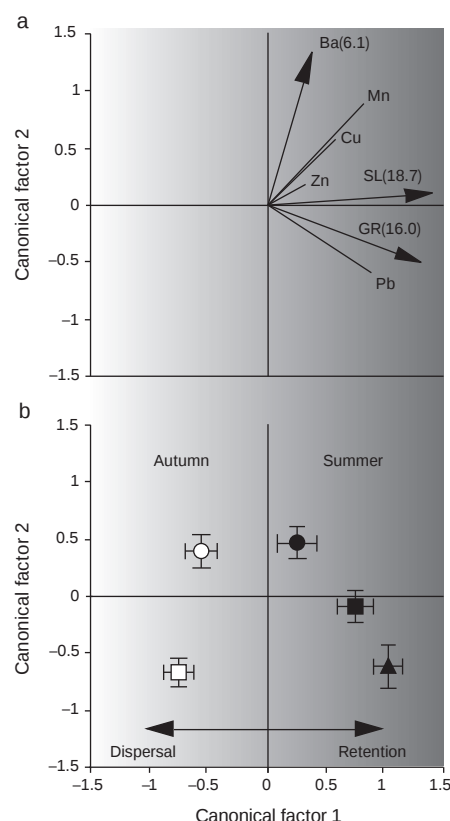


Figure 1 Results of the canonical discriminant analysis of larval source population signatures. Along the canonical factor 1 axis, predominantly elevated growth rates, larval size, and trace-element concentrations (retention signature) occur in larvae recruiting in summer relative to larvae recruiting in autumn. **a**, Vector plot depicting the strength and direction of the correlations between the dependent variables and the first two canonical factors³⁰. Vectors with arrows indicate endpoints beyond the scale of the figure (values in parentheses). **b**, Centroids (means) for each recruitment period. Bars = ± 1 s.e. See Table 1 legend for key to chemical symbols, and growth-rate and fish-size abbreviations. Filled circles, June; filled squares, July; filled triangles, August; open circles, September; open squares, October.

hypotheses, larvae with positive canonical factor 1 (CF1) scores were large, fast-growing individuals with otoliths enriched in trace elements (that is, retention characteristics) whereas larvae with negative CF1 scores were small, slow-growing individuals whose otoliths had low trace-element concentrations (that is, dispersal characteristics).

High proportions of larvae settling during the summer months were characterized by retention signatures (positive CF1 scores) in contrast to larvae settling in the autumn, which were characterized by dispersal signatures (negative CF1 scores). During the summer recruitment periods, there were strong differences among sites in the CF1 scores (Table 1), with 89% of the leeward-shore recruits with retention signatures compared to only 32% of the windward-shore recruits with such signatures. The variation in the prevalence of larvae with retention signatures had important repercussions for the intensity of recruitment to St Croix (Fig. 2a–c). Using the CF1 score (from Fig. 1) as an index of the intensity of local retention (positive values represent high retention and negative values represent low retention/high dispersal), we discovered pronounced differences between the leeward sites and the windward site in the relative importance of local retention of larvae to recruitment. At both leeward reefs (Butler Bay and Northstar, Fig. 2a, b), recruitment levels were positively correlated with CF1, indicating that the heavy summer recruitment events followed periods when large numbers of larvae tended to be locally retained (larvae had both high growth rates and high otolith trace-element concentrations).

Table 1 Comparisons of larval source population signatures among sites and among monthly recruitment events

Effect	MANOVA	Dependent variables (ANOVAs)						
	Pillai trace	Mn	Cu	Zn	Ba	Pb	GR	SL
Site	4.77†	3.02*	3.73*	0.36	8.46†	1.65	2.38	3.88*
Month	6.22†	3.18*	3.78†	2.04	8.45†	3.26*	18.76†	20.73†

Symbols used: Mn, manganese; Cu, copper; Zn, zinc; Ba, barium; Pb, lead; GR, somatic growth rate; SL, adjusted standard length. Trace-element concentrations were log transformed to conform to assumptions of the statistical tests. The MANOVA results indicate that significant overall differences existed among sites (first row) and months (second row). Significant ANOVA tests identify the important variables in the MANOVA models. Shown are *F*-ratios for tests based on differences in larval growth histories and otolith elemental composition among sites and months. Not shown are the pairwise multivariate comparisons (Hotelling's *T*²) among sites and months; all were significant at the appropriate Bonferroni adjusted significance level except for the July versus August comparison. Hotelling's *T*² comparisons among sites within season (summer versus autumn) indicated that significant differences occurred among sites in summer (June and July) but not in autumn (September and October). The untransformed means (ranges) for elements (in $\mu\text{mol mol}^{-1}$ Ca) were: Mn, 3.83 (1.19–29.38); Cu, 15.74 (4.11–59.98); Zn, 32.81 (4.46–126.77); Ba, 3.70 (2.05–8.22); Pb, 1.95 (1.06–6.95), and for growth histories: GR, 0.23 (0.18–0.28) mm d⁻¹; SL, 12.05 (9.95–13.9) mm.

**p* ≤ 0.05, †*p* ≤ 0.007 (Bonferroni adjusted α = 0.05/7), ‡*p* ≤ 0.001.

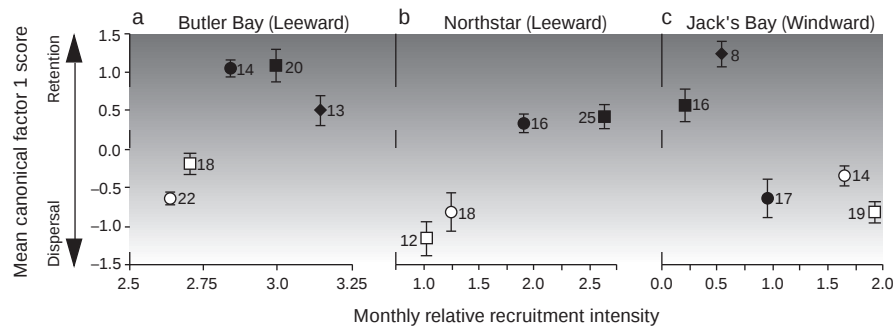


Figure 2 Relationship between larval retention and the intensity of recruitment. These data are evidence for high recruitment levels of bluehead wrasse, *Thalassoma bifasciatum*, coinciding with periods of greater larval retention (positive CF1 score) at leeward sites (**a**, **b**) in contrast to a windward site (**c**) when high recruitment levels occurred during periods dominated by dispersing larvae (negative CF1 score). Shown are the mean monthly CF1 scores and relative recruitment intensity for each site (bars show ± 1 s.e., see Fig. 1 for monthly symbol definitions). Relative recruitment is defined as [site average monthly recruit density (no. of fish per m^2)]/[island average monthly recruit density (no. of fish per m^2)]. Values adjacent to symbols are sample sizes (note: no data were available for the August Northstar recruit collection due to poor sample preserva-

tion). Under the assumption that each larva is an independent recorder of environmental conditions, the relationship between CF1 and recruitment intensity was strong and significant at all sites (Butler Bay, $r = 0.462$, $p < 0.001$; Northstar, $r = 0.546$, $p < 0.001$; Jack's Bay, $r = -0.476$, $p < 0.001$). A more conservative approach, using monthly mean CF1 values for each site, yields stronger Pearson product-moment correlation coefficients but these are not significant due to small sample size (Butler Bay, $r = 0.686$, $p = 0.201$; Northstar, $r = 0.922$, $p = 0.078$; Jack's Bay, $r = -0.784$, $p = 0.116$). Mean CF1 and recruitment were significantly correlated when data from the two leeward sites were pooled ($r = 0.752$, $p = 0.019$).

In contrast, recruitment levels at Jack's Bay (Fig. 2c) were inversely correlated with CF1, indicating that the elevated windward recruitment events in the autumn were related to larval transport through offshore water (larvae had both low larval growth rates and low otolith trace-element concentrations). In addition, the presence of positive CF1 scores during low summer recruitment events at Jack's Bay indicates that a small proportion of retention-signature larvae can settle to windward reefs.

Our findings suggest that recruitment to island populations, in particular to leeward reefs, may often result from the retention of locally spawned larvae. Local retention, compared with dispersal through oceanic waters, may be the more pervasive and adaptive larval dispersal strategy for many coastal marine organisms with planktonic larvae, given the more favourable environmental conditions for larval growth and survival and the higher probability of encountering suitable adult habitat at the end of larval development.

For bluehead wrasse populations on St Croix, recruitment is not simply a result of settlement from a common larval source. Instead, larvae settling to windward and leeward reefs spent significant amounts of time in different environments, indicating that processes affecting larval dispersal not only operate at intermediate spatial scales (>100 km; between-island dispersal) but also at much smaller spatial scales (<10 km; within-island dispersal). Temporal differences in larval growth rates and otolith trace-element signatures among recruitment periods indicate that reefs probably receive a mixture of locally retained and dispersing larvae over time. However, variation in recruitment intensity indicates that most of the recruits to a site may come from a primary source: locally retained larval populations for leeward reefs, and dispersing larval populations for windward reefs. Given the high rate of settlement on leeward reefs ($>70\%$ of the annual recruitment to St Croix), these findings suggest that a significant proportion of recruitment to St Croix results from the retention of locally produced larvae. While 'retention signatures' could occur in larvae produced on other islands that disperse quickly and spend the majority of larval development accumulating in coastal waters before settlement, we tested for such an effect by estimating the minimum number of days a larva would need to spend in coastal waters to produce the high otolith trace-metal concentrations observed in this study (see Methods: analysis of retention signatures). On the basis of these conservative estimates, a minimum of 50% of the recruits with retention signatures (positive CF1 scores) had spent their entire larval life in coastal waters.

The results we report here have implications for evaluating the efficacy of marine reserves, as reserve location may strongly determine the extent to which reserves can function to protect self-sustaining populations or enhance fisheries by supplying recruits to exploited populations. □

Methods

Larval growth rate and size

For all recruits ($n = 232$), we used standard length (SL in mm) and planktonic larval duration determined from otolith increment counts to calculate average somatic growth rates (mm d^{-1}). To remove post-settlement (PS) effects on growth, SL was regressed against PS age (number of increments after the settlement mark), and the residuals of the model were used in subsequent analyses (mean PS age (range): 3 (0–7) days).

Otolith trace-element analyses

On the basis of previous research^{25,26} and the coastal water chemistry of St Croix²⁷, we chose the following elements for otolith analyses: manganese, copper, zinc, barium and lead. Both sagittal otoliths from each recruit were: (1) crushed and immersed for 72 h in a 5% NaOCl solution to remove organic contaminants, (2) rinsed three times at 80 °C in 18 MΩ water to remove residual NaOCl, (3) transferred to acid-leached vials and (4) dissolved to ~ 1 mM Ca to reduce matrix suppression effects. We measured element concentrations on a VG Plasma Quad 2+ Inductively-coupled plasma mass spectrometer using a combination of internal standardization and external calibration techniques. Measurement error was $<5\%$ relative standard deviation for all elements. We standardized concentrations to molar calcium concentration to remove the effects of variable sample loss during sample cleaning. Since differences in otolith elemental signatures between larvae developing in water masses with different trace-element concentrations could be confounded by somatic growth rate (or temperature)²⁸, otolith precipitation rate²⁹, or post-settlement otolith growth, we tested for such effects on otolith elemental composition of bluehead wrasse recruits. Results of multiple analysis of covariance (MANCOVA) and subsequent univariate analysis of covariance (ANCOVA) tests for the five trace elements analysed produced nonsignificant effects ($p > 0.05$) for all three covariates.

Canonical discriminant analysis

We used canonical discriminant analysis (CDA) to determine which dependent variables in the multiple analysis of variance (MANOVA) model were most important for separating groups. For our model, months were maximally separated by weighting more heavily the otolith concentrations of barium, manganese, copper and lead, and both growth rate and standard length of each larva in the calculation of the canonical factors. In addition, these variables (Ba, Mn, Cu, Pb, GR and SL) were all positively correlated with canonical factor 1 (which explained 53% of the between-group variance), indicating that differences between months were consistent for all variables. That is, larvae with elevated otolith trace-element concentrations were also larger and grew faster. Additional complexity in the differences among recruitment periods was revealed by CF2 (which explained an additional 31% of the between-group variance). Along this axis, recruits showed strong differences in barium concentrations among months. This probably reflects the temporal cycling of barium which has a nutrient-type distribution¹⁴. Measurements of the monthly average barium concentration in sea water around St Croix in 1997 (S.E.S., unpublished data) show a strong correlation with CF2 ($r = 0.974$),

suggesting that the additional complexity in the data captured in CF2 reflects this temporal variability in barium.

Analysis of retention signatures

We estimated the minimum number of days a larva must spend in coastal waters in order to recruit with a retention signature. This calculation was based on the following: (1) A maximum 3-fold difference in seawater lead concentration between St Croix coastal and oceanic waters measured in 1997 and a $\text{Pb}/\text{Ca} = 0.8 \mu\text{mol mol}^{-1}$ concentration ratio in otoliths of reef fish larvae collected from oceanic waters in 1997 (S.E.S., unpublished data) as well as in otoliths from bluehead wrasse recruits with dispersal signatures collected in 1992, and (2) the assumption that trace elements are incorporated into otoliths in proportion to seawater concentrations. First, we calculated the percentage of the total otolith mass contained within each day of growth for a larva that spent 45 days in the plankton (average age at settlement for a bluehead wrasse larva). For each day spent in oceanic waters, the fraction of the otolith that grew incorporated $0.8 \mu\text{mol Pb per mol Ca}$. For each day spent in coastal water, the fraction of the otolith that grew incorporated $2.4 \mu\text{mol Pb per mol Ca}$. Assuming that days spent in oceanic waters occur before movement into coastal waters, we then calculated lead concentrations for larvae which spent zero days up to 45 days in coastal waters (the whole larval life). Finally, we compared the distribution of predicted lead concentrations to the distribution of observed lead concentrations. On the basis of this comparison, during the high summer recruitment pulses to leeward reefs, more than 50% of the larvae with retention signatures spent their entire larval duration in coastal waters.

Received 4 June; accepted 30 September 1999.

- Hjort, J. Fluctuations in the great fisheries of northern Europe viewed in the light of biological research. *Rapp. P.-V. Réun. Cons. Perm. Int. Explor. Mer* **20**, 1–228 (1914).
- Doherty, P. J. & Williams, D. McB. The replenishment of coral reef fish populations. *Oceanogr. Mar. Biol. Annu. Rev.* **26**, 487–551 (1988).
- Sale, P. F. In *The Ecology of Fishes on Coral Reefs* (ed. Sale, P. F.) 564–598 (Academic, San Diego, 1991).
- Caley, M. J. *et al.* Recruitment and the local dynamics of open marine populations. *Annu. Rev. Ecol. Syst.* **27**, 477–500 (1996).
- Roughgarden, J., Iwasa, Y. & Baxter, C. Demographic theory for an open marine population with space-limited recruitment. *Ecology* **66**, 54–67 (1985).
- Warner, R. R. & Hughes, T. P. In *Proc. 6th Int. Coral Reef Symp.* (eds Choat, J. H. *et al.*) Vol. 1 149–155 (Coral Reef Symposium Executive Committee, Townsville, 1988).
- Sammarco, P. W. & Andrews, J. C. Localized dispersal and recruitment in Great Barrier Reef corals: the helix experiment. *Science* **239**, 1422–1424 (1988).
- Wolanski, E. & Hamner, W. Topographically-controlled fronts in the ocean, and their influence on the distribution of organisms. *Science* **241**, 177–181 (1988).
- Neilson, J. D. & Perry, R. I. Diel vertical migrations of marine fishes: an obligate or facultative process? *Adv. Mar. Biol.* **26**, 115–168 (1990).
- Stobutzki, I. C. & Bellwood, D. R. Sustained swimming abilities of the late pelagic stages of coral reef fishes. *Mar. Ecol. Prog. Ser.* **149**, 35–41 (1997).
- Denman, K. L. & Powell, T. M. Effects of physical processes on planktonic ecosystems in the coastal ocean. *Oceanogr. Mar. Biol. Annu. Rev.* **22**, 125–168 (1984).
- Heywood, J., Barton, E. D. & Simpson, J. H. The effects of flow disturbance by an oceanic island. *J. Mar. Res.* **48**, 55–73 (1990).
- Rissik, D., Suthers, I. M. & Taggart, C. T. Enhanced zooplankton abundance in the lee of an isolated reef in the south Coral Sea: the role of flow disturbance. *J. Plankton Res.* **19**, 1347–1368 (1997).
- Bruland, K. W. In *Chemical Oceanography* Vol. 8 (eds Riley, J. P. & Chester, R.) 157–220 (Academic, San Diego, 1983).
- Campana, S. E. & Neilson, J. D. Microstructure of fish otoliths. *Can. J. Fish. Aquat. Sci.* **42**, 1014–1032 (1985).
- Mugiya, Y., Hakomori, T. & Hatsutori, K. Trace metal incorporation into otoliths and scales in the goldfish, *Carassius auratus*. *Comp. Biochem. Physiol. C* **99**, 327–331 (1991).
- Campana, S. E., Gagné, J. A. & McLaren, J. W. Elemental fingerprinting of fish otoliths using ID-ICPMS. *Mar. Ecol. Prog. Ser.* **122**, 115–120 (1995).
- Gillanders, B. M. & Kingsford, M. J. Elements in otoliths may elucidate the contribution of estuarine recruitment to sustaining coastal reef populations of a temperate reef fish. *Mar. Ecol. Prog. Ser.* **141**, 13–20 (1996).
- Thorrold, S. R., Jones, C. M. & Campana, S. E. Response of otolith microchemistry to environmental variations experienced by larval and juvenile Atlantic croaker (*Micropogonias undulatus*). *Limnol. Oceanogr.* **42**, 102–111 (1997).
- Caselle, J. E. & Warner, R. R. Variability in recruitment of coral reef fishes: the importance of habitat at two spatial scales. *Ecology* **77**, 2488–2504 (1996).
- Gordon, A. L. Circulation of the Caribbean Sea. *J. Geophys. Res.* **72**, 6207–6223 (1967).
- Corredor, J. E., Morell, J. & Mendez, A. Dissolved nitrogen, phytoplankton biomass and island mass effects in the northeastern Caribbean Sea. *Carib. J. Sci.* **20**, 129–132 (1984).
- Boehlert, G. W., Watson, W. & Sun, L. C. Horizontal and vertical distributions of larval fishes around an isolated oceanic island in the tropical Pacific. *Deep-Sea Res.* **39**, 439–466 (1992).
- Cowen, R. K. & Castro, L. R. Relation of coral reef fish larval distributions to island scale circulation around Barbados, West Indies. *Bull. Mar. Sci.* **54**, 228–244 (1994).
- Shen, G. T. & Boyle, E. A. Determination of lead, cadmium, and other trace metals in annually-banded corals. *Chem. Geol.* **67**, 47–62 (1988).
- Lea, D. W., Shen, G. T. & Boyle, E. A. Coralline barium records temporal variability in equatorial Pacific upwelling. *Nature* **340**, 373–376 (1989).
- U.S. Environmental Protection Agency U.S. Virgin Islands Ambient Monitoring Survey Sept. 15–23, 1992 (Environ. Services Div. Region 2, Edison, NJ, 1995).
- Hoff, G. R. & Fuiman, L. A. Morphometry and composition of red drum otoliths: changes associated with temperature, somatic growth rate, and age. *Comp. Biochem. Physiol. A* **106**, 209–219 (1993).
- Morse, J. W. & Bender, M. L. Partition coefficients in calcite: examination of factors influencing the validity of experimental results and their application to natural systems. *Chem. Geol.* **82**, 265–277 (1990).
- Overall, J. E. & Klett, C. J. *Applied Multivariate Analysis* (McGraw-Hill, New York, 1972).

Acknowledgements

We thank M. Sheehy for assistance with sample collection, L. Linn, P. Martin and T. Mashiotto for assistance with ICP-MS analyses, and N. Barbee, S. Cooper, S. Gaines, K. Lafferty and J. Shima for comments on the manuscript. This work was supported by the US NOAA-National Undersea Research Program, the US National Science Foundation, and the David and Lucille Packard Foundation.

Correspondence and requests for materials should be addressed to S.E.S. (e-mail: swearer@lifesci.ucsb.edu).

Self-recruitment in a coral reef fish population

G. P. Jones, M. J. Milicich, M. J. Emslie & C. Lunow

School of Marine Biology and Aquaculture, James Cook University, Townsville, 4811 Queensland, Australia

The question of how far the larvae of marine organisms disperse is fundamental to an understanding of their population dynamics^{1–3}, the management of exploited species^{4,5} and the conservation of marine biodiversity^{6,7}. It is generally assumed that larvae disperse away from their natal population so that local populations operate as ‘open’ systems, driven by recruitment of larvae from other sub-populations⁸. However, this assumption has never been critically tested. Here we show for the first time that juveniles from a coral reef fish population can return to their natal reef. We marked otoliths (ear bones) of over 10 million developing embryos of the damselfish, *Pomacentrus amboinensis*, at Lizard Island (Great Barrier Reef). Subsequently, from an examination of 5,000 juveniles settling at the same location, we found 15 marked individuals. On the basis of an estimate of the proportion of embryos marked (0.5–2%), as many as 15–60% of juveniles may be returning to their natal population (self-recruitment). We challenge the assumption that long-distance dispersal is the norm for reef fish populations.

The majority of marine organisms (~70%) have a so-called ‘dispersive’ planktonic larval stage in their early life history before recruiting into adult habitat⁹. However, just how far these juveniles normally travel from their natal area is an unanswered question in marine biology^{10,11}. The potential for dispersal by ocean currents is so great that self-recruitment within local populations has been considered extremely unlikely. However, over the last few years information has been accumulating that suggests a degree of self-recruitment. This includes the unexpected genetic subdivision of some marine populations^{12,13}, the persistence of endemic species with pelagic larvae on small isolated islands¹⁴, the persistence of new populations established from marine introductions¹⁵, the persistence of populations with no upcurrent source¹⁶ and new information on the behaviour of larvae in the vicinity of reefs^{17–19}. However, none of this information can provide direct and irrefutable evidence for the existence and importance of self-recruitment as a mechanism for population maintenance. Here we test for the first time whether any juveniles of a marine fish species return to their parent population, if they do, we assess the proportion of the population that is self-recruiting.

The only way to unequivocally answer these questions is to mark larvae so that they can be traced from their origin to their ultimate destination. Mark/recapture techniques to assess larval dispersal have not been attempted in the marine environment. We developed a technique of marking the otoliths of developing embryos of the small coral reef damselfish, *Pomacentrus amboinensis*, a common species on the northern Great Barrier Reef. It allows them to be marked en masse in their natural habitat, and to hatch and disperse

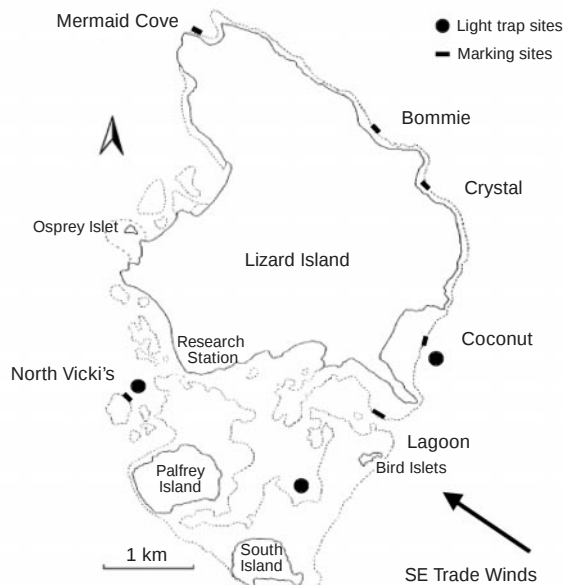


Figure 1 Map of Lizard Island on the northern Great Barrier Reef, showing the location of the six 150-m stretches of reef edge where all embryos of *Pomacentrus amboinensis* were marked over a three-month period (October–December 1994). Light traps were placed at three sites to collect incoming larvae ready to settle onto the reef, with four light traps at the windward site, two at the lagoon site and two at the back reef site. Dotted lines indicate the reef.

with minimal interference. We adapted the use of the fluorescent compound tetracycline, which has been widely used to mark otoliths in fishes and validate age rings laid down²⁰, and more recently, to mark the hatchery-reared larvae of freshwater and estuarine fishes^{21–24}. Preliminary laboratory experiments showed that immersion of embryos in tetracycline resulted in clear fluorescent marks at all stages of development (even prior to otolith formation). Tetracycline produced a signature core of yellow fluorescence and there are no natural sources known to produce fluorescence of this type in fish otoliths. Laboratory rearing of larvae 28 days after hatching showed that the mark persisted through larval development and metamorphosis.

We were able to use this technique to mark large numbers in the field because, like many marine species, female *P. amboinensis* lay their eggs on the substratum and they readily spawn onto artificial surfaces. Embryos spawned onto PVC tiles could be marked *in situ* with 100% success by discharging tetracycline into plastic bags temporarily containing the tile. A comparison of hatching success of marked and unmarked embryos, with appropriate controls for the disturbance effect, showed no significant reduction in survivorship (Embryo mortality = $16.1 (\pm 2.1)$ standard error)% for unmarked clutches and $17.9 (\pm 3.9)$ s.e. for marked clutches).

Over a three-month period encompassing most of the 1994 breeding season, we marked an estimated 10 million embryos at sites around Lizard Island, on the northern Great Barrier Reef (Fig. 1). A total of 1,926 clutches were marked, with an average clutch size of $4,696 (\pm 535)$ s.e.). On the basis of the summed area of clutches marked over the study ($70,120 \text{ cm}^2$) and a mean egg density of $142.9 (\pm 4.4)$ s.e.) we estimate that $10,020,148 (\pm 308,529)$ s.e.) embryos were marked.

We collected 7,327 juveniles from light traps, deployed daily at Lizard Island over a three-month period coinciding with marking, with an appropriate lag phase for the length of larval life. The otoliths of a random selection of 5,000 of these were dissected out and examined using a fluorescence microscope. A characteristic circular tetracycline mark near the centre indicated those fish that were marked during the embryo stage. A total of 15 (0.3% of juveniles examined) had sagittal otoliths with a clear yellow, tetra-

Box 1

Estimate of self-recruitment

To calculate the level of self-recruitment (S) in the population, we used the following mark/recapture formula:

$$\%S = O/E \times 100$$

Here O is the observed number of tags in captured fish (15), and E is the expected number of tags assumed complete self-recruitment. E is calculated as:

$$E = R \times M_e/T_e$$

Here R is the number of captured fish (5000), and M_e/T_e is the proportion of the total number of embryos produced at Lizard Island that were marked. Based on aerial maps, we conservatively estimated that M_e/T_e ranged between 0.5 and 2%. Therefore:

$$\%S = 15\text{--}60$$

cycline mark, which unequivocally demonstrates for the first time that coral reef fish larvae are capable of returning to their natal reef. These 15 recaptures are extremely significant as, given the potential for oceanic dilution and mortality, recapturing any marked larvae was previously considered to be impossible. Marked juveniles came from light traps at all three locations, with most coming in during new-moon peaks in recruitment.

Sufficient recaptures were obtained to provide an approximate estimate of the proportion of the total number of recruits that returned to Lizard Island (Box 1). This required an estimate of the proportion of the total number of embryos produced at Lizard Island over the 3-month period that were marked. Based on a number of assumptions (Box 1), the estimate of the proportion of eggs marked ranged from 0.5% to 2% of the total population. Given the range in the estimate of the total population, our estimates of self-recruitment range from 15 to 60% (Box 1). Both extremes suggest that self-recruitment is a much more significant component of the replenishment process than was previously thought.

The level of self-recruitment we have calculated is far greater than would be predicted by most models based on passive larval dispersal. It requires that a considerable proportion of the individuals in a population have the ability to maintain themselves in the vicinity of their natal reef. If this is also typical of other species, substantial recruitment to natal populations must be incorporated into population models and management strategies. Given some self-recruitment, population models may need to be developed that relate recruitment to parental stock on a smaller scale than previously considered. Patterns of recruitment ought not, perhaps, to be viewed as a completely independent determinant of local population dynamics or age structure as the literature suggests^{2,3,8}. Management strategies based solely on source–sink relationships⁷ need to examine the effects of varying the proportions of self-recruitment and downstream dispersal. We argue that models that over-emphasize downstream dispersal will not predict the long-term behaviour of populations and suggest that both domestic and regional initiatives will contribute to the successful management of local populations. □

Methods

Embryos were marked over a 90-day period (October–December 1994) at six sites around Lizard Island (Fig. 1). Each site consisted of a 150-m stretch of reef edge and lower reef slope, the primary breeding habitat where male *P. amboinensis* defend nest sites. PVC tiles were placed next to all breeding males in each area (~70 males per site). All males switched to defending the tiles as nest sites and all females laid their eggs on the tiles in these areas. Every two days, all tiles containing developing embryos were placed in plastic bags and the embryos immersed in a tetracycline solution ($400 \mu\text{M}$ per litre) for a 1 hour period. As embryos take 4–5 days to hatch, all embryos were marked at least once prior to hatching. Therefore it is assumed that all embryos hatching along each 150-m stretch of reef were marked over the three-month period. The number of embryos marked was calculated by recording the number and area of all new clutches on the tiles (mapped onto acetate

sheets) and by estimating the average egg density per clutch (measured from 50 clutches using macro-photography).

Fully developed larvae of *P. amboinensis* settle onto reefs at night after a restricted period of 18–21 days in the plankton^{25,26}. At this time they are easily identified and strongly attracted to light, so they can readily be collected at night using light traps²⁷. Four light traps were placed at a windward site, two in the lagoon and two at a back-reef location between 21 October and 21 January, in order to sample juveniles returning to different reef zones (Fig. 1).

The calculation of the proportion of the total number of embryos produced at Lizard Island that were marked is based on the size of the six areas in which all the embryos were marked, as a proportion of an estimate of the total suitable breeding habitat at Lizard Island. The area of suitable breeding habitat was assumed to be proportional to the length of the reef-edge habitat (a total of 900 m for marking sites). The length of all reef-edge habitat was measured from an aerial photograph (scale 1 : 7500) using a digitizing tablet. The upper limit excluded all adjacent reef habitats within 500 m that were separated from Lizard Island by stretches of sand, and the lower limit included such reefs.

The calculation of M_e/T_e (Box 1) using relative breeding area assumes that: (1) All embryos produced at the six marking sites over the three months were actually marked. This can be substantiated as all females laid their eggs on tiles and all tiles with eggs were immersed every two days. Also, the pilot study showed that immersion in tetracycline for 1 h is 100% successful. (2) The relative sizes of the areas in which marking was and was not carried out reflect the difference in egg production between the two areas. There were no significant differences in mean clutch size or female density where tiles were present and where they were absent. Mean clutch size = $4.969 (\pm 535 \text{ s.e.})$ at the six marking sites and $5.237 (\pm 625 \text{ s.e.})$ at three non-marking sites; mean female density = $17.5 (\pm 3.5 \text{ s.e.})$ per 100 m² at the six marking-sites and $18.4 (\pm 3.1 \text{ s.e.})$ at three non-marking sites. (3) There was no difference in the mortality rates and/or recapture probabilities of marked and unmarked embryos. Our pilot study showed that at least until hatching, survival rates of marked and unmarked embryos were not statistically different. In the three-month study only 12 PVC tiles over-turned between visits. This only affected five marked clutches. A further five marked clutches appeared to have been decimated by benthic predators. These were not included in the final calculations.

Received 25 March; accepted 25 October 1999.

1. Sinclair, M. *Marine Populations: an Essay on Population Regulation and Speciation* (Washington Sea Grant, Seattle, 1988).
2. Roughgarden, J., Gaines, S. D. & Pacala, S. W. in *Organization of Communities* (eds Gee, J. H. R. & Giller, P. S.) 491–518 (Blackwell, London, 1987).
3. Underwood, A. J. & Fairweather, P. G. Supply-side ecology and benthic marine assemblages. *Trends Ecol. Evol.* **4**, 16–20 (1989).
4. Roberts, C. M. Connectivity and management of Caribbean coral reefs. *Science* **278**, 1454–1457 (1997).
5. Roberts, C. M. Ecological advice for the global fisheries crisis. *Trends Ecol. Evol.* **12**, 35–38 (1997).
6. Fairweather, P. G. Implications of 'supply-side ecology' for environmental assessment and management. *Trends Ecol. Evol.* **6**, 60–63.
7. Roberts, C. M. & Hawkins, J. P. Extinction risk in the sea. *Trends Ecol. Evol.* **14**, 241–246.
8. Caley, M. J., Carr, M. H., Hixon, M. A., Hughes, T. P., Jones, G. P. & Menge, B. A. Recruitment and the local dynamics of open marine populations. *Annu. Rev. Ecol. Syst.* **27**, 477–500 (1996).
9. Thorson, G. Reproductive and larval ecology of marine bottom invertebrates. *Biol. Rev. Cambridge Phil. Soc.* **25**, 1–45 (1950).
10. Palumbi, S. R. Genetic divergence, reproductive isolation, and marine speciation. *Annu. Rev. Ecol. Syst.* **25**, 547–572 (1994).
11. Shulman, M. J. What can population genetics tell us about dispersal and biogeographic history of coral-reef fishes? *Aust. J. Ecol.* **23**, 216–225 (1998).
12. Bell, L. J., Moyer, J. T. & Numachi, K. Morphological and genetic variation in Japanese populations of the anemonefish *Amphiprion clarkii*. *Mar. Biol.* **72**, 99–108 (1982).
13. Planes, S. Genetic differentiation in relation to restricted larval dispersal of the convict surgeonfish *Acanthurus triostegus* in French Polynesia. *Mar. Ecol. (Prog. Ser.)* **98**, 237–246 (1993).
14. Hourigan, T. F. & Reese, E. S. Mid-ocean isolation and the evolution of Hawaiian reef fishes. *Trends Ecol. Evol.* **2**, 187–191 (1987).
15. Baltz, D. M. Introduced fishes in marine systems and inland seas. *Biol. Conserv.* **56**, 151–177 (1991).
16. Schultz, E. T. & Cowen, R. K. Recruitment of coral-reef fishes to Bermuda: local retention or long-distance transport? *Mar. Ecol. (Prog. Ser.)* **109**, 15–28 (1994).
17. Doherty, P. J. & Carleton, J. M. The distribution and abundance of pelagic juvenile fish near Grub Reef, central Great Barrier Reef. *Proc. 8th Int. Coral Reef Symposium* **2**, 1155–1160 (1997).
18. Stobutzki, I. C. & Bellwood, D. R. Nocturnal orientation to reefs by late pelagic stage coral reef fishes. *Coral Reefs* **17**, 103–110 (1998).
19. Leis, J. M., Sweatman, H. P. A. & Reader, S. E. What the pelagic stages of coral reef fishes are doing out in blue water: daytime field observations of larval behavioural capabilities. *Mar. Freshwater Res.* **47**, 401–411 (1996).
20. Campana, S. E., Neilson, J. D. Microstructure of fish otoliths. *Can. J. Fish. Aquat. Sci.* **42**, 1014–1032.
21. Secor, D. H., White, M. G. & Dean, J. M. Immersion marking of larval and juvenile hatchery-produced striped bass with oxytetracycline. *Trans. Am. Fish. Soc.* **120**, 261–266 (1991).
22. Secor, D. H. & Houde, E. D. in *Recent Developments in Fish Otolith Research* (eds Secor, D. D., Dean, J. M. & Campana, S. E.) 423–444 (University of South Carolina Press, Columbia, 1995).
23. Tsukamoto, K. Otolith tagging of ayu embryo with fluorescent substances. *Nippon Suisan Gakkaishi* **54**, 1289–1295 (1988).
24. Reinert, T. R., Wallin, J., Griffin, M. C., Conroy, M. J. & Van Den Avyle, M. J. Long-term retention and detection of oxytetracycline marks applied to hatchery-reared larval striped bass, *Morone saxatilis*. *Can. J. Fish. Aquat. Sci.* **55**, 539–543 (1998).
25. Meekan, M. G., Milicich, M. J. & Doherty, P. J. Larval production drives temporal patterns of larval supply and recruitment of a coral reef damselfish. *Mar. Ecol. (Prog. Ser.)* **93**, 217–225 (1993).
26. Kerrigan, B. A. Temporal patterns in size and condition at settlement in two tropical reef fishes (Pomacentridae: *Pomacentrus amboinensis* and *P. nagasakiensis*). *Mar. Ecol. (Prog. Ser.)* **135**, 27–41 (1996).

27. Doherty, P. J. Light-traps: selective but useful devices for quantifying the distributions and abundances of larval fishes. *Bull. Mar. Sci.* **41**, 423–431 (1987).

Acknowledgements

We thank I. Keay and B. Kerrigan for assistance in the field and laboratory, and M. McCormick and P. Munday for editorial comments. This work was supported by the Australian Research Council, Large Grant Scheme, James Cook University and the Australian Museum's Lizard Island Research Station.

Correspondence and requests for materials should be addressed to G.P.J. (e-mail: geoffrey.jones@jcu.edu.au).

Regulation of lifespan by sensory perception in *Caenorhabditis elegans*

Javier Apfeld & Cynthia Kenyon

Department of Biochemistry and Biophysics, University of California at San Francisco, California 94143-0448, USA

Caenorhabditis elegans senses environmental signals through ciliated sensory neurons located primarily in sensory organs in the head and tail. Cilia function as sensory receptors, and mutants with defective sensory cilia have impaired sensory perception^{1,2}. Cilia are membrane-bound microtubule-based structures and in *C. elegans* are only found at the dendritic endings of sensory neurons³. Here we show that mutations that cause defects in sensory cilia or their support cells, or in sensory signal transduction, extend lifespan. Our findings imply that sensory perception regulates the lifespan of this animal, and suggest that in nature, its lifespan may be regulated by environmental cues.

We determined the lifespans of eleven mutants (in nine genes) with various defects in sensory cilia, including the absence of cilia (*daf-19*), deletion of the middle and distal segments (*che-2*, *che-13*, *osm-1*, *osm-5* and *osm-6*) and reduced or irregular cilia segments (*che-3*, *che-11* and *daf-10*). One of these genes, *osm-6*, has been cloned. *osm-6* is expressed only in ciliated neurons⁴, and encodes a homologue of the *Chlamydomonas reinhardtii* intraflagellar transport particle, required for assembly of the flagellum, a cilium-like structure⁵. We found that all of these mutants were long lived (Fig. 1; Table 1). Thus, the function of sensory cilia is required for *C. elegans* to age normally.

Sensory cilia are present in 60 of the 302 neurons present in the *C. elegans* hermaphrodite³. The amphids, a pair of lateral sensilla in the head, are the principal sensory organs. Each amphid contains the ciliated endings of twelve sensory neurons, plus a sheath and a socket cell, which form a pore to the exterior. The phasmids are similar but smaller sensory organs in the tail. Four other classes of cuticular sensilla are also located in the head. We determined the lifespans of five mutants in four genes thought to affect only a subset of sensilla. *daf-6(e1377)* mutants, in which the amphid and phasmid pores are closed because of a defect in the sheath cells⁶, were long lived (Fig. 1; Table 1). *mec-8(e398)* mutants, in which amphid sensory cilia partially fail to fasciculate¹, as well as two *osm-3* mutants, in which only the cilia of the phasmid neurons and 16 out of the 24 amphid neurons are defective¹, were also long lived (Fig. 1; Table 1). *osm-3* encodes a kinesin-like protein that is expressed in the 20 ciliated sensory neurons that are affected in the mutant, and in 6 other ciliated sensory neurons that are unaffected in the mutant⁷. The remaining mutant, a *mec-1* mutant that has a relatively weak defect in amphid cilia fasciculation¹, was not long lived (Fig. 1; Table 1). Finally, to test the involvement of the amphids directly, we ablated the two amphid sheath cells with a laser microbeam and found that this treatment also extended lifespan

(Fig. 1; Table 1). Together, these findings suggest that at least some of the neurons whose activity influences lifespan are amphid sensory neurons.

In vertebrate olfactory and visual systems, cyclic nucleotide-gated channels are required to transduce receptor activity into electrical activity⁸. In *C. elegans*, the cyclic nucleotide-gated channel α -subunit TAX-4 and β -subunit TAX-2 function at the ciliated endings of a subset of amphid neurons to mediate several sensory behaviours^{9,10}. *tax-2* and *tax-4* are thought to function directly in sensory transduction⁹; mutations in these genes do not affect the structure of the sensory cilia¹⁰. We tested two *tax-4* alleles and found that both extended lifespan (Fig. 1; Table 1). We also tested three

tax-2 alleles, and found that they had little or no effect on lifespan (Fig. 1; Table 1). TAX-4 is able to form functional channels in the absence of TAX-2 (ref. 11). These findings suggest that a TAX-4-dependent channel that does not require TAX-2 may operate within sensory cilia to regulate lifespan.

We then considered whether cilium structure mutants had any other obvious defects that might influence lifespan. We analysed mutants in five genes, *che-2*, *daf-6*, *daf-10*, *osm-3* and *osm-5*. The feeding behaviours of these mutants were normal. The percentage of animals pumping (a pharyngeal behaviour that draws bacteria into the mouth and intestine¹²) was similar to wild type, as was the rate of pumping in individual animals (Table 2). The mutants were less likely to stray from the bacterial lawns (their food) than were wild-type animals (Table 2); this preference for food may be caused by

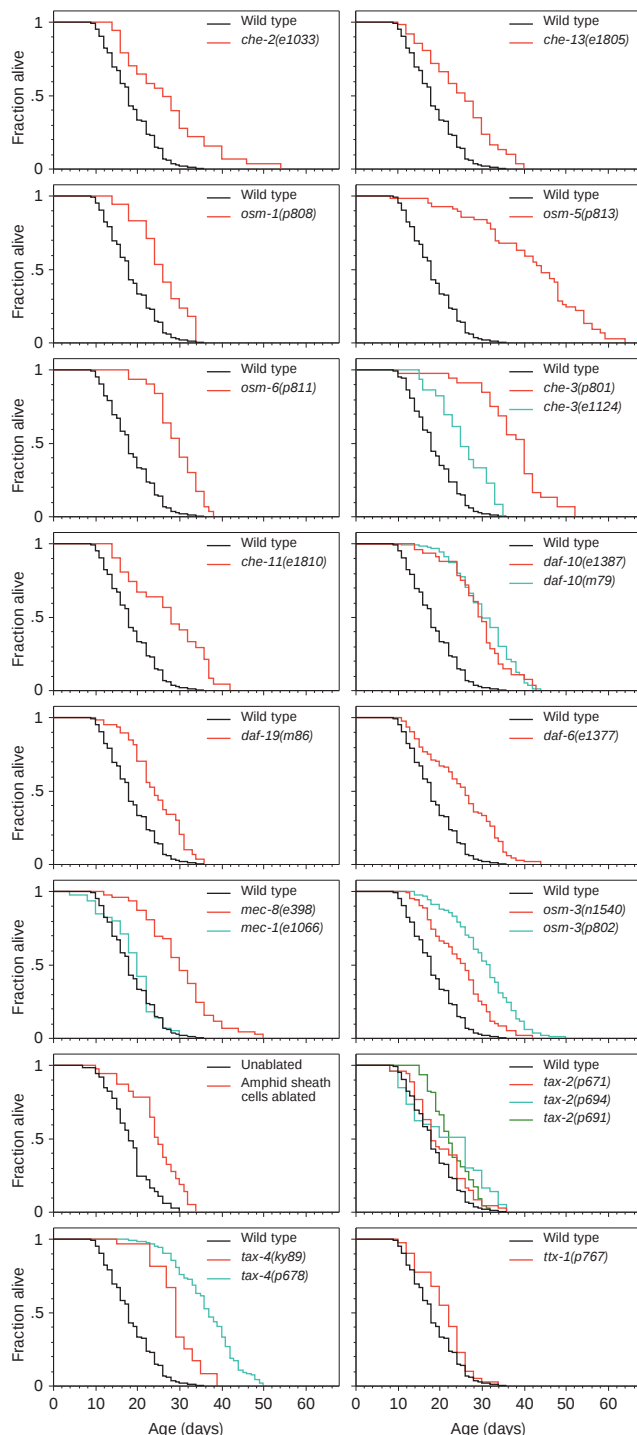


Figure 1 Animals with defects in sensory cilia and sensory transduction live longer than wild type. The fraction of animals remaining alive is plotted against animal age.

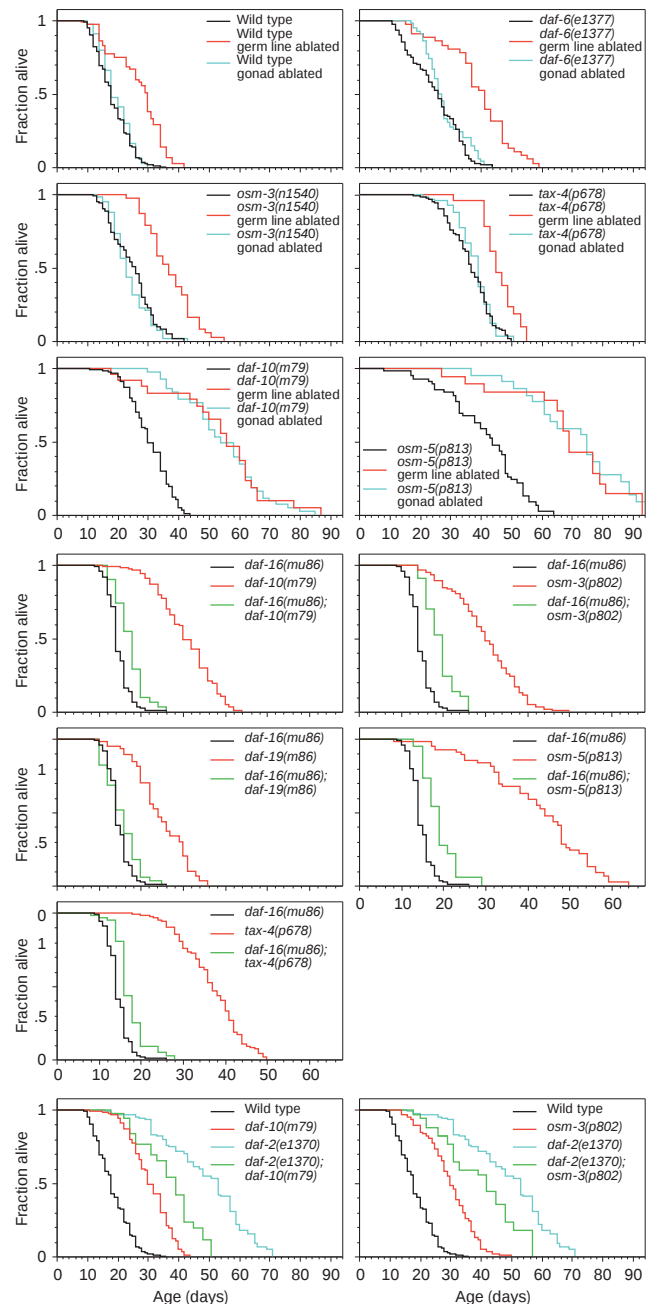


Figure 2 Survival curves of germline- and gonad-ablated animals, and double mutants. The fraction of animals remaining alive is plotted against animal age. The deaths of the last two surviving *osm-5(p813)* gonad-ablated animals are not shown. These animals died when they were 95 and 121 days old.

their altered sensory ability². The mutants had a normal appearance and length of postembryonic development, and they had either normal or slightly elevated brood sizes (Table 2). Because these mutants appeared normal in all of these respects, it is possible that their longevity is caused directly by altered sensory perception.

In *C. elegans*, signals from the reproductive system influence lifespan¹³; the germ line produces signals that shorten lifespan, and the somatic gonad appears to produce counterbalancing signals that lengthen lifespan¹³. We ablated the germline precursors, Z2 and Z3, in five mutants, *daf-6*, *daf-10*, *osm-3*, *osm-5* and *tax-4*, and found that their lifespans were extended (Fig. 2; Table 3). Thus, germline signalling appeared normal. We then removed the entire gonad by killing the somatic gonad precursors, Z1 and Z4. This treatment prevents the development of the germ line, as well as the somatic gonad, and has no effect on wild-type lifespan^{13,14}. Three of the mutants, *daf-6*, *osm-3* and *tax-4*, behaved like wild type; that is, killing the entire gonad had no effect on their lifespans (Fig. 2; Table 3). In contrast, in *daf-10* and *osm-5* mutants, killing the entire gonad lengthened lifespan to the same extent as did killing only the

germ line (Fig. 2; Table 3). Thus, in these mutants, somatic gonad signalling appeared to be silenced. *daf-10* and *osm-5* mutations affect most of the ciliated sensory neurons, whereas *daf-6*, *osm-3* and *tax-4* mutations affect only subsets of these neurons¹⁹. Therefore, it is possible that somatic gonad signalling requires the activity of sensory neurons that are affected in *daf-10* and *osm-5* but not in *daf-6*, *osm-3* or *tax-4* mutants. Alternatively, impaired function of a large number of neurons could be required to silence somatic gonad signalling.

The lifespan of *C. elegans* is thought to be regulated hormonally. Mutations in *daf-2*, which encodes an insulin/IGF-1 receptor homologue, and in downstream signalling components, cause animals to live more than twice as long as wild type^{14,15} (reviewed in ref. 16). This longevity requires DAF-16, a forkhead-family transcription factor^{14,17–20}. *daf-16* also acts in pathways that appear to regulate lifespan independently of *daf-2*, such as the germline signalling pathway¹³.

To investigate the role of *daf-16*, we constructed double mutants between a *daf-16* null allele, *mu86* (ref. 20), and mutations in five genes, *osm-3*, *osm-5*, *daf-10*, *daf-19* and *tax-4*. Loss of *daf-16*

Table 1 Adult lifespans at 20 °C

Strain/treatment	Mean $\pm$ s.e.m. (days)	75th percentile (days)*	Number of animals that died/Total†	P
Wild type	18.8 $\pm$ 0.3	22	347/531 (11)	
<i>che-2(e1033)</i>	26.8 $\pm$ 1.7	32	33/48 (1)	<0.0001‡
<i>che-3(p801)</i>	37.5 $\pm$ 1.4	42	31/47 (1)	<0.0001‡
<i>che-3(e1124)</i>	25.7 $\pm$ 1.0	31	38/60 (1)	<0.0001‡
<i>che-11(e1810)</i>	27.3 $\pm$ 1.6	36	27/49 (1)	<0.0001‡
<i>che-13(e1805)</i>	25.0 $\pm$ 1.3	30	36/95 (2)	0.0005, 0.0007§
<i>daf-6(e1377)</i>	25.1 $\pm$ 1.0	33	75/101 (2)	<0.0001‡
<i>daf-10(e1387)</i>	29.4 $\pm$ 1.2	34	30/100 (2)	<0.0001‡
<i>daf-10(m79)</i>	30.7 $\pm$ 0.5	36	137/198 (4)	<0.0001‡
<i>daf-19(mu86)</i>	24.5 $\pm$ 1.0	30	35/126 (2)	0.0003, 0.0005§
<i>mec-1(e1066)</i>	19.0 $\pm$ 0.4	24	45/49 (1)	0.45
<i>mec-8(e398)</i>	30.0 $\pm$ 1.2	34	46/50 (1)	<0.0001‡
<i>osm-1(p808)</i>	25.8 $\pm$ 1.4	30	17/54 (1)	<0.0001‡
<i>osm-3(n1540)</i>	24.8 $\pm$ 0.8	30	67/101 (2)	<0.0001‡
<i>osm-3(p802)</i>	30.6 $\pm$ 0.6	36	133/248 (5)	<0.0001‡
<i>osm-5(p813)</i>	41.6 $\pm$ 1.7	50	46/146 (3)	<0.0001‡
<i>osm-6(p811)</i>	29.4 $\pm$ 1.0	34	30/48 (1)	<0.0001‡
<i>txx-1(p767)**</i>	21.0 $\pm$ 0.9	24	40/49 (1)	0.15
Unablated	18.3 $\pm$ 0.8	20	41/60 (1)	
Amphid sheath cells ablated	24.5 $\pm$ 1.2	29	23/49 (1)	<0.0001¶
<i>tax-4(p678)</i>	36.6 $\pm$ 0.7	42	105/145 (3)	<0.0001‡
<i>tax-4(ky89)</i>	29.1 $\pm$ 1.4	33	13/49 (1)	<0.0001‡
<i>tax-2(p671)</i>	20.1 $\pm$ 0.9	24	50/61 (1)	0.19
<i>tax-2(p691)</i>	22.7 $\pm$ 0.5	27	100/120 (1)	0.34
<i>tax-2(p694)</i>	21.6 $\pm$ 1.4	30	44/54 (1)	0.01
<i>daf-16(mu86)</i>	14.5 $\pm$ 0.2	16	169/263 (4)	
<i>daf-16(mu86); daf-10(m79)</i>	17.4 $\pm$ 0.6	20	31/48 (1)	<0.0001#
<i>daf-16(mu86); daf-19(mu86)</i>	15.3 $\pm$ 0.6	18	36/60 (1)	<0.0001, 0.32‡
<i>daf-16(mu86); osm-3(p802)</i>	19.2 $\pm$ 0.6	20	29/47 (1)	<0.0001‡
<i>daf-16(mu86); osm-5(p813)</i>	18.7 $\pm$ 0.6	21	36/60 (1)	<0.0001‡
<i>daf-16(mu86); tax-4(p678)</i>	17.3 $\pm$ 0.5	20	45/84 (1)	<0.0001‡
<i>daf-2(e1370)</i>	49.2 $\pm$ 1.5	59	82/103 (1)	
<i>daf-2(e1370); daf-10(m79)</i>	37.2 $\pm$ 1.8	42	27/49 (1)	<0.0001**
<i>daf-2(e1370); osm-3(p802)</i>	40.0 $\pm$ 2.1	48	34/48 (1)	<0.0001**

* The 75th percentile is the age when the fraction of animals alive reaches 0.25.

† The total number of observations equals the number of animals that died plus the number censored. The number of independent times the lifespan was determined is in parenthesis. Animals that crawled off the plate, exploded or bagged were censored at the time of the event. This step incorporated those worms into the data set until the censor date²⁷, and was necessary to avoid the loss of information; for example, if a 50-day-old animal crawls off the plate, it is important to include that information in the data set, as that animal was long lived. Control and experimental animals were cultured in parallel. Many experiments were performed more than once. The logrank (Mantel-Cox) test was used to test the hypothesis that the survival functions among groups were equal. P values were calculated for individual experiments, each consisting of control and experimental animals examined at the same time. We show the cumulative statistics in figures and tables because strains behaved similarly in different experiments. Below, the letters A–K refer to the 11 different experiments we performed; in each experiment, multiple strains were examined in parallel. The P values comparing mean lifespans of the same strain in different experiments were P = 0.40 for wild type (mean lifespans of animals tested in each of 11 individual experiments: A, 17.2 $\pm$ 0.9; B, 18.9 $\pm$ 0.8; C, 19.7 $\pm$ 0.8; D, 18.4 $\pm$ 0.9; E, 17.9 $\pm$ 0.9; F, 19.0 $\pm$ 1.0; G, 17.8 $\pm$ 0.8; H, 19.3 $\pm$ 1.0; I, 20.5 $\pm$ 1.1; J, 18.4 $\pm$ 1.1; K, 20.0 $\pm$ 1.0); P = 0.20 for *che-13* (means: B, 23.7 $\pm$ 1.6; J, 26.2 $\pm$ 2.2); P = 0.81 for *daf-6* (means: C, 24.3 $\pm$ 1.4; J, 25.8 $\pm$ 1.4); P = 0.27 for *daf-10* (means: E, 27.5 $\pm$ 2.1; J, 30.8 $\pm$ 1.3); P = 0.33 for *daf-10* (m79) (means: A, 31.1 $\pm$ 1.0; C, 30.0 $\pm$ 1.1; G, 29.3 $\pm$ 1.3; I, 32.4 $\pm$ 0.9); P = 0.92 for *daf-19* (means: E, 24.3 $\pm$ 1.4; F, 24.5 $\pm$ 1.4); P = 0.02 for *osm-3* (n1540) (means: D, 26.8 $\pm$ 1.2; J, 22.7 $\pm$ 1.0); P = 0.04 for *osm-3* (p802) (means: A, 29.0 $\pm$ 1.4; C, 32.8 $\pm$ 1.7; G, 29.1 $\pm$ 1.6; I, 29.0 $\pm$ 1.2; J, 32.8 $\pm$ 1.1); P = 0.60 for *osm-5* (means: B, 40.9 $\pm$ 2.8; F, 41.7 $\pm$ 2.4; J, 44.9 $\pm$ 5.0); P = 0.52 for *tax-4* (p678) (means: I, 36.3 $\pm$ 1.6; J, 36.3 $\pm$ 1.1; K, 37.1 $\pm$ 1.0); and P = 0.38 for *daf-16* (means: C, 14.1 $\pm$ 0.2; E, 14.9 $\pm$ 0.4; F, 14.7 $\pm$ 0.5; J, 14.1 $\pm$ 0.4). Strains whose lifespans were determined only once were examined in the following experiments: A, *osm-1*; B, *che-2*; C, *daf-16*; *daf-10*; *daf-16*; *osm-3*; D, *che-3* (p801); *mec-1*; *mec-8*; E, *che-11*; *osm-6*; F, *daf-16*; *osm-5*; *daf-16*; *daf-19*; G, *daf-2*; *daf-2*; *daf-10*; *daf-2*; *osm-3*; J, *txx-1*; *tax-4* (ky89); *tax-2* (p671); *tax-2* (p691); *tax-2* (p694); *daf-16*; *tax-4*. The experiments in which the germline and gonad ablations were performed (Table 3) were I, *daf-10* and wild type; J, *daf-6*, *osm-3*, *osm-5*, *tax-4* and wild type (for wild type, only intact and germ-line-ablated animals were examined in this experiment). The total number of animals was at least 46 for each strain in each experiment.

‡ Compared with wild type control in all experiments.

§ Compared with wild-type control in two experiments.

|| Compared with wild-type control.

¶ Compared with unablated control.

Compared with cilium-structure mutant control and *daf-16* (mu86) control.

‡ Compared with *daf-19* (mu86) and *daf-16* (mu86) controls, respectively.

** Compared with cilium-structure mutant control and *daf-2* (e1370) control.

†† In addition, *txx-1* (p767) animals had lifespans that were not significantly different from those of wild type at 15 °C (P = 0.03) and 25 °C (P = 0.37).

Table 2 Phenotypic characterization of cilium-structure mutants

Genotype	Percentage of animals on the bacterial lawn*	Feeding rate (pumps per minute)†	Percentage of animals pumping*	Length of postembryonic development (h)‡	Brood size‡
Wild type	65% (312)	206 ± 12 (20)	100% (320)	44.8 ± 1.1 (18)	280 ± 18 (11)
<i>che-2(e1033)</i>	81% (110)	210 ± 12 (5)	100% (110)	44.1 ± 2.4 (37)	272 ± 36 (9)
<i>daf-6(e1377)</i>	90% (104)	210 ± 14 (5)	100% (100)	44.9 ± 1.3 (31)	327 ± 20 (12)
<i>daf-10(m79)</i>	85% (113)	204 ± 15 (5)	100% (340)	45.1 ± 1.7 (40)	306 ± 10 (12)
<i>osm-3(p802)</i>	80% (71)	208 ± 8 (5)	100% (180)	44.6 ± 1.0 (36)	294 ± 31 (12)
<i>osm-5(p813)</i>	83% (174)	210 ± 14 (5)	100% (245)	44.9 ± 1.5 (17)	321 ± 31 (12)

Fisher's exact tests were conducted to determine the probability that the observed differences in the percentage of animals on the bacterial lawn were due to chance. Unpaired *t*-tests were conducted to determine the probability that the observed differences in mean feeding rate, mean length of postembryonic development and mean brood size between two groups were due to chance. With two exceptions, no statistical significance was observed when comparing wild type with any group ($P > 0.1$). The measured mean brood sizes of *daf-6* ($P < 0.0001$), *daf-10* ($P = 0.0002$) and *osm-5* ($P = 0.0008$) were larger than wild type, and the fractions of *che-2* ($P = 0.0002$), *daf-6* ($P < 0.0001$), *daf-10* ($P < 0.0001$), *osm-3* ($P = 0.012$) and *osm-5* ($P < 0.0001$) mutants located on the bacterial lawn were greater than wild type.

* Number of one-day-old adult animals assayed is in parentheses.

† Mean ± s.d. Number of trials is in parentheses. Each trial represents the number of pumps observed in a one-day-old adult animal in 1 min.

‡ Mean ± s.d. Number of animals assayed is in parentheses.

markedly reduced but, in most cases, did not completely eliminate lifespan extension (Fig. 2; Table 1). Thus, most of this lifespan extension requires *daf-16* activity, but a fraction is *daf-16* independent. *daf-16* could act downstream of a sensory signal to regulate lifespan, or it could act in a parallel pathway to provide an activity that these animals simply require for their longevity.

We also determined the lifespans of *daf-2* double mutants. This method has been used to show that the PI3 kinase *age-1* is likely to function in a pathway with *daf-2*, as the double mutant has a lifespan that is similar to those of the single mutants¹⁸. In contrast, the longevity produced by germline ablation appears to be *daf-2* independent, because ablating the germ line in *daf-2* mutants causes an additional doubling of lifespan¹³. We constructed *daf-2(e1370)* double mutants with *daf-10* and *osm-3* mutations. The double mutants did not live longer than the *daf-2(e1370)* single mutant; instead, unexpectedly, their lifespans were slightly shorter than those of *daf-2(e1370)* (Fig. 2, Table 1). A trivial explanation, that the cilium-structure mutations somehow limit the potential lifespan of the animal, seems unlikely, as germline-ablated *daf-10* mutants lived much longer than *daf-2*; *daf-10* double mutants (Fig. 2, Table 3). Thus, we infer that the cilium-structure mutants do not act exclusively in a *daf-2*-independent pathway. To reconcile these findings, we propose the model that sensory neurons do act in a *daf-2*-dependent pathway, but that they also affect the animal in another way. Specifically, we suggest that an environmental signal perceived by sensory cilia regulates a DAF-2 ligand. In cilium-structure mutants, the level of this ligand falls and lifespan is

extended. In addition, we propose that the cilium-structure mutants produce a second, weaker signal that shortens lifespan. In this model, if a *daf-2* mutation completely mimics the lifespan-extending effect of the cilium-structure mutants, but does not completely block the effect of the lifespan-shortening signal, the double mutants would have a lifespan that is shorter than that of the *daf-2* single mutant.

Certain long-lived *daf-2* mutants, such as the tyrosine kinase-domain mutant *e1370* (ref. 15), also appear to lack somatic gonad signalling¹³. In contrast, the *daf-2(e1368)* mutant, which has an altered DAF-2 ligand-binding domain¹⁵, has normal somatic gonad signalling¹³. This suggests that DAF-2 can respond to two ligands, one of which is somatic-gonad dependent (and can still bind to the *e1368* receptor), and one of which is not (and cannot activate either mutant receptor)¹³. Perhaps the cilium-structure mutations that block somatic gonad signalling affect both DAF-2 ligands, whereas the cilium-structure mutations that do not block somatic gonad signalling affect only the gonad-independent ligand.

Unlike weak loss of *daf-2* function, which increases adult lifespan, more severe loss of *daf-2* function causes juvenile animals to become dauer larvae²¹, a stress-resistant, developmentally arrested larval form normally induced by food limitation, high temperature and crowding²². The ability of *daf-2* mutants to become dauer larvae requires the activity of *daf-16* (ref. 21). Sensory cilia are also involved in the regulation of dauer larva formation^{1,6}, and cilium-structure and *tax-4* mutations are known to interact genetically with other mutations in genes that affect dauer larva formation^{6,21,23,24}.

Table 3 Adult lifespans of germline-ablated and gonad-ablated animals at 20 °C

Strain/treatment	Mean ± s.e.m. (days)	75th percentile (days)	Number of animals that died/total	<i>P</i> *
Wild type	18.4 ± 0.3	22	347/531 (11)	
Wild type germline ablated	26.9 ± 1.1†	34	49/112 (2)	<0.0001
Wild type gonad ablated	19.5 ± 0.8	24	44/47 (1)	0.30
<i>daf-6(e1377)</i>	25.1 ± 1.0	33	75/101 (2)	
<i>daf-6(e1377)</i> germline ablated	39.0 ± 1.7	47	39/66 (1)	<0.0001
<i>daf-6(e1377)</i> gonad ablated	28.1 ± 0.9	33	57/81 (1)	0.20
<i>osm-3(n1540)‡</i>	24.8 ± 0.8	30	67/101 (2)	
<i>osm-3(n1540)</i> germline ablated	37.2 ± 1.2	43	57/62 (1)	<0.0001
<i>osm-3(n1540)</i> gonad ablated	24.1 ± 0.8	27	38/54 (1)	0.27
<i>tax-4(p678)</i>	36.6 ± 0.7	42	105/145 (3)	
<i>tax-4(p678)</i> germline ablated	46.1 ± 1.2	49	22/51 (1)	<0.0001
<i>tax-4(p678)</i> gonad ablated	38.1 ± 0.7	43	60/67 (1)	0.52
<i>daf-10(m79)</i>	30.7 ± 0.5	36	137/198 (4)	
<i>daf-10(m79)</i> germline ablated	53.9 ± 3.5	62	22/47 (1)	<0.0001
<i>daf-10(m79)</i> gonad ablated	54.6 ± 2.0	64	43/60 (1)	<0.0001
<i>osm-5(p813)</i>	41.9 ± 1.8	50	48/146 (3)	
<i>osm-5(p813)</i> germline ablated	68.6 ± 4.3	86	15/35 (1)	<0.0001
<i>osm-5(p813)</i> gonad ablated	72.8 ± 4.0	79	22/48 (1)	<0.0001

Germ line ablated: the germ line precursors Z2 and Z3 were ablated. Gonad Ablated: the somatic gonad precursors Z1 and Z4 were ablated. Killing Z1 and Z4 also blocks the development of the germline. The experiments in which these ablations were performed are listed in the legend of Table 1.

* *P* values were calculated for individual experiments, each consisting of unablated controls and experimental animals examined at the same time.

† Means were: experiment I, 25.8 ± 1.6 and experiment J, 28.2 ± 1.6, $P = 0.41$.

‡ We also ablated the gonads of *osm-3(p802)*; these animals had lifespans that were similar to those of unablated controls ($P = 0.69$, mean = 29.0 ± 1.2, dead/total = 28/47). We were unable to estimate the lifespan of *osm-3(p802)* germ-line-ablated animals, as most of the animals were censored due to explosions.

Table 4 Percentage of animals that formed dauers at 27 °C

Genotype	% Dauer 27 °C*
Wild type	5 ± 2 (22, n = 611)
<i>age-1(hx546)</i>	97 ± 2 (22, n = 349)
Sensory cilia mutants	
<i>che-2(e1033)</i>	97 ± 3 (2, n = 29)
<i>che-3(p801)</i>	31 ± 13 (2, n = 58)
<i>che-3(e1124)</i>	89 ± 7 (2, n = 49)
<i>che-11(e1810)</i>	96 ± 0 (2, n = 52)
<i>che-13(e1805)</i>	98 ± 2 (2, n = 56)
<i>daf-10(m79)</i>	90 ± 2 (4, n = 103)
<i>osm-1(p808)</i>	94 ± 6 (2, n = 55)
<i>osm-3(n1540)</i>	50 ± 3 (3, n = 69)
<i>osm-5(p813)</i>	100 ± 0 (4, n = 104)
<i>osm-6(p811)</i>	100 ± 0 (2, n = 57)
Sheath cells mutant	
<i>daf-6(e1377)</i>	0 ± 0 (2, n = 65)
Cilia fasciculation mutants	
<i>mec-1(e1066)</i>	10 ± 6 (2, n = 44)
<i>mec-8(e398)</i>	2 ± 2 (2, n = 48)
Sensory transduction mutants	
<i>tax-4(p678)</i>	77 ± 7 (4, n = 119)
<i>tax-4(ky89)</i>	100 ± 0 (4, n = 120)
<i>daf-16</i> mutants	
<i>daf-16(mu86)</i>	0 ± 0 (6, n = 170)
<i>daf-16(mu86); daf-10(m79)</i>	0 ± 0 (2, n = 44)
<i>daf-16(mu86); osm-5(p813)</i>	0 ± 0 (2, n = 55)
<i>daf-16(mu86); tax-4(p678)</i>	0 ± 0 (10, n = 300)

* Mean ± s.e.m. Number of trials and total number of animals tested (n) shown in parentheses.

Some mutations have been shown to cause dauer formation at relatively high temperature, 27 °C (refs 16, 25), including cilium-structure mutants (J. Thomas, personal communication). We confirmed this, and found that *tax-4* mutants also formed dauer larvae constitutively at 27 °C. This dauer-formation phenotype, like that of *daf-2*, was *daf-16* dependent in each of three mutants, *osm-5(p813)*, *daf-10(m79)* and *tax-4(p678)* that we tested (Table 4). This suggests that, during dauer formation, sensory input regulates the activity of *daf-16*, possibly by changing the level of *daf-2* activity. This 27 °C dauer phenotype was not observed in mutants with defects in sheath cells or cilia fasciculation (Table 4); thus, the roles of sensory structures in dauer formation and adult lifespan can be uncoupled.

In summary, our findings suggest that perception of an environmental signal may regulate the lifespan of *C. elegans*. What might this signal be? The amphid neurons mediate responses to temperature, volatile and soluble odorants, and mechanical stimuli. It is unlikely that the signal is only temperature, because *ttx-1* mutants, which lack the sensory endings of the amphid finger cells that mediate thermotaxis¹, have normal lifespans (Fig. 1; Table 1). In addition, the amphid finger cells are not affected by *osm-3* mutations¹, which do affect lifespan. The signal affecting lifespan may be a substance that the animal can smell or taste. For example, it may be a pheromone such as dauer pheromone, which reflects population density²², or a compound that originates from organic material and reflects food availability.

Our findings indicate that the ageing process of *C. elegans* is quite plastic and can be affected by signals from sources that would seem to be particularly relevant to its survival and reproductive success: its environment and its reproductive system. Moreover, sensory perception may affect reproductive signalling. Under laboratory conditions, the germ lines and somatic gonads of normal animals appear to produce counterbalancing signals¹³. In this situation, perturbations that retard the development or activity of the germ line but not the somatic gonad might lengthen lifespan, thus, possibly increasing the chances that the animal is still youthful enough to bear progeny when its germ line matures. Perhaps other environmental conditions mimic *daf-10* and *osm-5* mutations, in which germline signals are no longer counterbalanced by somatic gonad signals. Under these conditions, the ageing process could become much more sensitive to general perturbations affecting reproduction. □

Methods

Lifespan analysis

Lifespan assays were conducted at 20 °C as described^{13,26}; at the L4 moult, animals were transferred to plates containing 16 µM 5-fluoro-2'-deoxyuridine (FUDR, Sigma), which kills their progeny as embryos. Control experiments indicated that FUDR does not significantly affect lifespan. We used the L4 moult as *t* = 0 for lifespan analysis. Strains were grown at 20 °C for at least two generations before lifespan determination. A major cause of censoring was explosion, which generally occurred around day 10 of adulthood. In general, the pattern of censoring was not different among strains, except for *daf-19* and *osm-5* mutants, which were more likely to crawl off the plate. In addition, we were unable to estimate the lifespan of the *che-14(e1960)* mutants, because almost all animals exploded. We used Statview 5.0.1 (SAS) software to carry out statistical analysis, and to determine means and percentiles. Censored animals were incorporated into the data set as described²⁷.

Laser ablations

L1 larvae were mounted on 2%-agarose pads containing 1–4 µM NaN₃ as an anaesthetic within an hour of hatching. The amphid sheath cells (AMshL and AMshR) or gonad precursors were killed with a laser microbeam as described¹³. Unablated control animals were treated identically but no cells were killed. In separate controls, the AMshL and AMshR cell deaths were confirmed by the absence of DiI (1,1'-diiododecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate, Molecular Probes) dye filling in amphid neurons⁴.

Strain construction

In double-mutant construction, cilium-structure mutations were assayed by the inability of the amphid and phasmid neurons to fill with the dye DiI. *daf-2(e1370)* and *tax-4(p678)* were assayed by the presence of dauers at 25 °C and 27 °C, respectively. *daf-16(mu86)* and *daf-16(+)* were distinguished by PCR using allele-specific primers²⁰. *daf-16(mu86)* is not Dyf and did not suppress the Dyf defect of *daf-10(m79)*, *daf-19(m86)*, *osm-3(p802)* or *osm-5(p813)* mutants.

Other assays

For dauer assays, animals born at 20 °C were transferred onto a new well-seeded plate as eggs and incubated at 27 °C for 48 h. All other assays were conducted at 20 °C. To measure brood size, two (or in some cases one) L4 animals were cultured together and transferred to new plates every day or every 12 h, and their progeny counted. A set of reserve animals was cultured and transferred in parallel. In the case of *osm-5* (but not the other mutants we examined), animals whose progeny we were counting sometimes crawled off the plates. When this happened, the plate was replaced with another containing age-matched reserve animals, and their subsequent progeny were then counted. We used *F*-tests to compare the variance of the mean brood sizes and found no statistical difference when comparing wild type with any group (*P* > 0.05 in all cases). Pharyngeal pumping was assayed using a dissecting microscope. To determine the length of postembryonic development, larvae that hatched within an hour were monitored every 2 h as they neared adulthood and were scored as adults when they had undergone the final moult and a vulva could be observed.

Received 6 September; accepted 12 October 1999.

- Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986).
- Peckol, E. L., Zallen, J. A., Yarrow, J. C. & Bargmann, C. I. Sensory activity affects sensory axon development in *C. elegans*. *Development* **126**, 1891–1902 (1999).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *C. elegans*. *Phil. Trans R. Soc. Lond. B* **314**, 1–340 (1986).
- Collet, J., Spike, C. A., Lundquist, E. A., Shaw, J. E. & Herman, R. K. Analysis of *osm-6*, a gene that affects sensory cilium structure and sensory neuron function in *Caenorhabditis elegans*. *Genetics* **148**, 187–200 (1998).
- Cole, D. G. *et al.* Chlamydomonas kinesin-II-dependent intraflagellar transport (IFT): IFT particles contain proteins required for ciliary assembly in *Caenorhabditis elegans* sensory neurons. *J. Cell Biol.* **141**, 993–1008 (1998).
- Albert, P., Brown, S. & Riddle, D. Sensory control of dauer larva formation in *Caenorhabditis elegans*. *J. Comp. Neurol.* **198**, 435–451 (1981).
- Tabish, M., Siddiqui, Z. K., Nishikawa, K. & Siddiqui, S. S. Exclusive expression of *C. elegans osm-3* kinesin gene in chemosensory neurons open to the external environment. *J. Mol. Biol.* **247**, 377–389 (1995).
- Zagotta, W. N. & Siegelbaum, S. A. Structure and function of cyclic nucleotide-gated channels. *Annu. Rev. Neurosci.* **19**, 235–263 (1996).
- Komatsu, H., Mori, I., Rhee, J. S., Akaike, N. & Ohshima, Y. Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. *Neuron* **17**, 707–718 (1996).
- Coburn, C. M. & Bargmann, C. I. A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans*. *Neuron* **17**, 695–706 (1996).
- Komatsu, H. *et al.* Functional reconstitution of a heteromeric cyclic nucleotide-gated channel of *Caenorhabditis elegans* in cultured cells. *Brain Res.* **821**, 160–168 (1999).
- Avery, L. & Horvitz, H. R. Pharyngeal pumping continues after laser killing of the pharyngeal nervous system of *C. elegans*. *Neuron* **3**, 473–485 (1989).
- Hsin, H. & Kenyon, C. Signals from the reproductive system regulate the lifespan of *C. elegans*. *Nature* **399**, 362–366 (1999).
- Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).

15. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
16. Paradis, S., Ailion, M., Toker, A., Thomas, J. H. & Ruvkun, G. A PDK1 homolog is necessary and sufficient to transduce AGE-1 PI3 kinase signals that regulate diapause in *Caenorhabditis elegans*. *Genes Dev.* **13**, 1438–1452 (1999).
17. Larsen, P. L., Albert, P. S. & Riddle, D. L. Genes that regulate both development and longevity in *Caenorhabditis elegans*. *Genetics* **139**, 1567–1583 (1995).
18. Dorman, J. B., Albinder, B., Shroyer, T. & Kenyon, C. The *age-1* and *daf-2* genes function in a common pathway to control the lifespan of *Caenorhabditis elegans*. *Genetics* **141**, 1399–1406 (1995).
19. Ogg, S. *et al.* The fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
20. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. *daf-16*: an HNF-3-forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
21. Riddle, D., Swanson, M. & Albert, P. Interacting genes in nematode dauer larva formation. *Nature* **290**, 668–671 (1981).
22. Golden, J. & Riddle, D. The *Caenorhabditis elegans* dauer larva: developmental effects of pheromone, food, and temperature. *Dev. Biol.* **102**, 368–378 (1984).
23. Vowels, J. & Thomas, J. Genetic analysis of chemosensory control of dauer formation in *Caenorhabditis elegans*. *Genetics* **130**, 105–123 (1992).
24. Coburn, C. M., Mori, I., Ohshima, Y. & Bargmann, C. I. A cyclic nucleotide-gated channel inhibits sensory axon outgrowth in larval and adult *Caenorhabditis elegans*: a distinct pathway for maintenance of sensory axon structure. *Development* **125**, 249–258 (1998).
25. Ailion, M., Inoue, T., Weaver, C. L., Holdcraft, R. W. & Thomas, J. H. Neurosecretory control of aging in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **96**, 7394–7397 (1999).
26. Apfeld, J. & Kenyon, C. Cell nonautonomy of *C. elegans daf-2* function in the regulation of diapause and life span. *Cell* **95**, 199–210 (1998).
27. Lawless, J. F. *Models and Methods for Lifetime Data* (Wiley, New York, 1982).

Acknowledgements

We thank J. Whangbo, S. Alper, J. Alcedo, H. Hsin, K. Lin, L. Yang, Q. Ch'ng, A. Dillin, D. Garigan and other members of the Kenyon Lab, as well as members of Cori Bargmann's lab, for stimulating discussions. Some nematode strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH. J.A. was supported by an HHMI Predoctoral Fellowship. This work was supported by a grant from the NIA to C.K.

Correspondence and requests for materials should be addressed to C.K.
(e-mail: ckenyon@biochem.ucsf.edu).

Atomic scale movement of the voltage-sensing region in a potassium channel measured via spectroscopy

Albert Cha^{*}, Gregory E. Snyder[†], Paul R. Selvin[†] & Francisco Bezanilla^{*}

^{*} Department of Physiology and Department of Anesthesiology, UCLA School of Medicine, Los Angeles, California 90095, USA

[†] Department of Physics and Biophysics Center, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Voltage-gated ion channels are transmembrane proteins that are essential for nerve impulses and regulate ion flow across cell membranes in response to changes in membrane potential. They are made up of four homologous domains or subunits, each of which contains six transmembrane segments^{1,2}. Studies of potassium channels have shown that the second (S2) and fourth (S4) segments contain several charged residues, which sense changes in voltage and form part of the voltage sensor^{3–5}. Although these regions clearly undergo conformational changes in response to voltage^{6–10}, little is known about the nature of these changes because voltage-dependent distance changes have not been measured. Here we use lanthanide-based resonance energy transfer^{11,12} to measure distances between *Shaker* potassium channel subunits at specific residues. Voltage-dependent distance changes of up to 3.2 Å were measured at several sites near the S4 segment. These movements directly correlated with electrical measurements of the voltage sensor, establishing the link between physical changes and electrical charge movement. Measured distance changes suggest that the region associated with the S4 segment undergoes

a rotation and possible tilt, rather than a large transmembrane movement, in response to voltage. These results demonstrate the first *in situ* measurement of atomic scale movement in a transmembrane protein.

Lanthanide-based resonance energy transfer (LRET) is a modification of conventional fluorescence resonance energy transfer in which a long-lived lanthanide donor transfers energy in a distance-dependent manner to a conventional organic fluorescent acceptor^{11,12}. This technique has previously been used to measure Angstrom-scale conformational changes in proteins¹³. Here, specific sites in the channel were fluorescently labelled by substituting a cysteine for a particular residue and attaching cysteine-reactive compounds of either a donor, a terbium-chelate maleimide (TbM), or an acceptor, fluorescein maleimide (FM), to the same site in the four identical subunits of the channel (Fig. 1a; see also Methods).

Although this labelling leads to a heterogeneous population of channels with different numbers of donors and acceptors, associated problems are greatly minimized for three reasons. In LRET, only donor–acceptor pairs generate sensitized emission signals (that is, delayed emission of the acceptor after receiving energy from the donor), and channels that contain all donors or all acceptors do not contribute to the signal^{11,12}. Also, with four-fold symmetry of the channel^{14,15}, there are only two possible intersubunit distances: the distance between residues on neighbouring subunits and the distance between residues across the pore, which are related by the Pythagorean theorem (Fig. 1a). Finally, the labelling is done so that there is typically only one acceptor per channel, which can readily accept energy independently from multiple donors.

Intersubunit distances were examined at specific sites near S2, in the S3–S4 linker and S4, and near the channel pore (Fig. 2, Table 1). The intersubunit distances (Fig. 1c and Methods) were calculated using the time constants of acceptor-sensitized emission and donor emission without an acceptor. Donor lifetime in the absence of an acceptor was independent of voltage at all sites, indicating no significant change in the environment of the caged terbium (data not shown). The sensitized emission displayed two time constants at all sites measured, reflecting two donor–acceptor distances that follow a Pythagorean relationship (Fig. 1c). This verifies that the technique is measuring distances between contiguous subunits and across the pore. The distance measured at F425C across the pore (30 Å) corresponds well to the distance obtained between α -carbons for the homologous residue from the crystal structure of the bacterial analogue, KcsA (29 Å) (ref. 14). This confirms that the technique is capable of obtaining realistic estimates of distance, as the KcsA pore has been shown to be similar to the *Shaker* pore in its ability to bind specific inhibitors of *Shaker*¹⁶. However, LRET, like other energy transfer techniques, is better at measuring relative distances or changes in distance than absolute distances. Potential sources of error, such as anisotropy of the fluorophores and uncertainty in the fluorophore's position with respect to the labelled residue due to the linker, have been addressed (see Methods).

Next, voltage-dependent movements near the voltage-sensing regions were measured by determining intersubunit distances as a function of voltage. Residue S346C demonstrated a robust voltage-dependent change in energy transfer as measured by sensitized emission lifetimes 50 ms after changing the voltage (Fig. 3a). This change is probably consistent with a voltage-dependent movement of ~3.2 Å between S346C residues on contiguous subunits as the channel moves into the open state (Fig. 3b). The voltage-dependent movement of S346C also coincides with the gating charge movement for the same channels, as determined by parameters of a sequential three-state model fit²¹ to both values (Fig. 3c). The striking agreement between the measured distance changes and electrical charge movement strongly indicates that the observed movement of S346C residues is closely correlated with the movement of charged residues involved in channel gating.

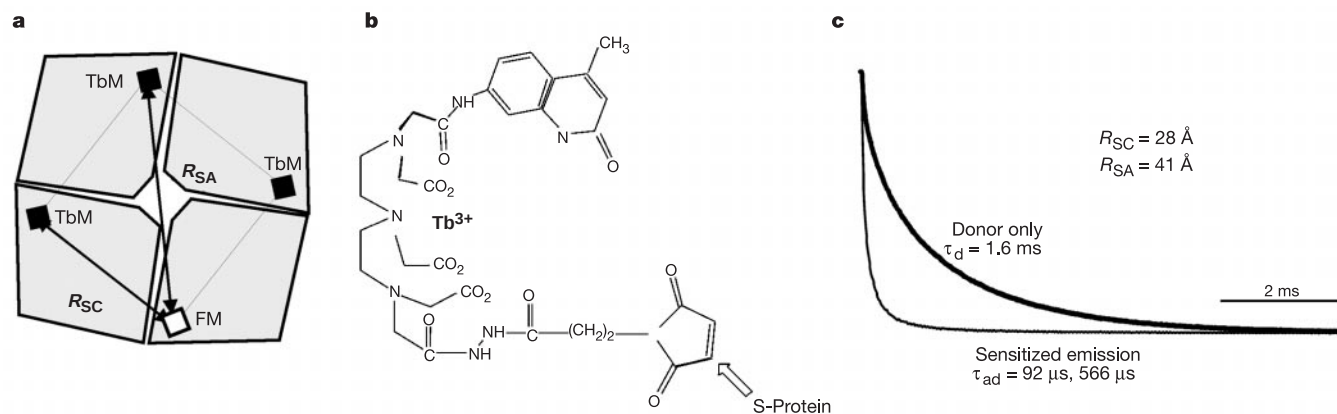


Figure 1 Labelling methodology and characteristics of ion channel. **a**, Labelling scheme of the channel's four identical subunits. About 70% of the channels that generate sensitized emission have three terbium-labelled subunits (TbM, closed squares) and one fluoroscein-labelled subunit (FM, open square). With four-fold symmetry, the two measured distances between donor and acceptor are R_{SC} (between contiguous subunits) and R_{SA} (across the pore), related by the Pythagorean equation $R_{SA}^2 = 2R_{SC}^2$. **b**, Structure of the donor terbium-maleimide chelate, with dimensions of approximately $13 \times 8 \times 8 \text{ \AA}$

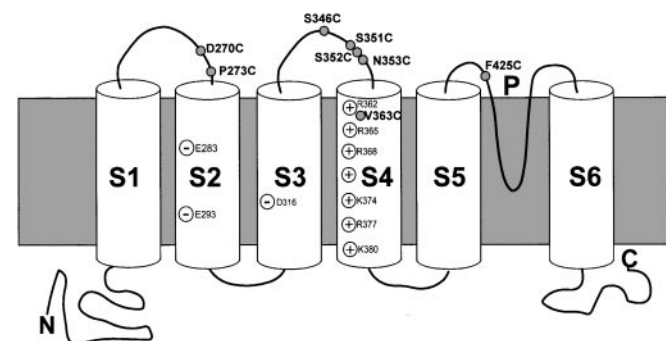


Figure 2 Transmembrane topology of the *Shaker* potassium channel. Cysteine-substituted sites examined with LRET measurements are denoted by grey circles and labelled in bold. Residues that constitute part of the voltage sensor are labelled and denoted as positively (+) and negatively (-) charged. P denotes the pore; C and N denote the C and N termini, respectively.

Table 1 Intersubunit distances and voltage-dependent changes measured using LRET

Location	Site	R_{SC} (Å)	R_{SA} (Å)	R_{SA} from R_{SC} (Å)	Standard error R_{SC}, R_{SA}	Anisotropy constraints (+%, -%)
S1-S2 loop	D270C	31	45	43	0.5, 0.6 $n = 9$	+6%, -3.8%
	P273C	32	47	46	0.7, 1.1 $n = 9$	+5.5%, -3.4%
S3-S4 loop	S346C	26-30	37-42	37-42	0.5, 0.7 $n = 5$	+4.5%, -2.7%
	S351C	28	41-42	40	0.4, 1.1 $n = 3$	+6.0%, -3.9%
	S352C	29	40	41	0.5, 1.6 $n = 13$	+5.3%, -3.3%
	N353C	29	42-44	41	0.6, 1.3 $n = 5$	+5.8%, -3.7%
S4	V363C	32	45	45	0.5, 0.6 $n = 7$	+5.4%, -3.3%
Pore	F425C		30	29	0.5 $n = 6$	

Shown are the measured distances between subunits (R_{SC}) and across the pore (R_{SA}) with standard error (n = number of oocytes) for sites throughout the channel. The estimated R_{SA} is calculated from the measured R_{SC} . The distance ranges given for S346C, S351C, and N353C indicate voltage-dependent distance changes from the closed (in bold) to the open state. The maximum error attributable to changes in the orientation factor κ^2 (last column) reflect the most extreme error possible, with a standard error of <0.2% (see Methods).

(ref. 24). S-protein represents the ion channel cysteine residue that attaches to the maleimide ring. **c**, Excited-state lifetimes used to deduce distances at site S346C. The terbium donor-only emission (thick trace) and fluorescein-sensitized emission (thin trace) are fitted to exponential decay functions, and the two corresponding time constants are used to compute two different distances (see Methods). Assuming that the shorter distance is R_{SC} , R_{SA} can be predicted from the measured R_{SC} (40 Å) and compared to the measured value (41 Å).

Energy transfer changes can arise not only from distance changes but also from a rotation of the acceptor or from spectral changes in the donor or acceptor. However, the magnitude of the energy transfer and distance changes at S346C is larger than could be attributed to a rotation of the acceptor, which could cause a worst-case apparent change of $\sim 2 \text{ \AA}$ (+4.5% or -2.7% of the 26-29 Å distance between subunits at this site) (Table 1). The actual error is likely to be considerably less (see Methods). Similarly, a 50% change in the spectral overlap, which would be necessary to account for the energy transfer changes, is inconsistent with the voltage-dependent acceptor intensity change of <2% measured at this site.

It is conceivable that the donor or acceptor attached to S346C undergoes movement without an underlying protein conformational change owing to a change in the local potential gradient. However, the accessibility of site S346C to charged fluorophores is independent of voltage, and there is very little voltage-dependent change in fluorescence intensity, anisotropy, or emission spectra from fluorophores placed at the site (data not shown). Furthermore, if the electric field affected the position of the dyes, one would expect to see distance changes at site V363C which is expected to be in the strongest electric field of all the residues tested, because it lies between two of the gating charges in S4. However, V363C shows no intersubunit change in distance with voltage. Also, a voltage-dependent movement of the dyes, independent of the protein, would probably not correlate well with gating charge movement. Thus, it is likely that the measured distance changes are caused by protein conformational changes, and not caused by movement of the dye(s) alone.

It is surprising that movements (at S346C) can be detected at residues as many as 16 amino acids from the critical charged residues in S4, and that these movements show a voltage dependence of movement mirroring that of the electrical charge displacement. This result implies that the S3-S4 linker has a stiff secondary structure that allows movements of S4 to be faithfully transmitted throughout the linker. According to our measured distances in the S3-S4 linker in the closed state, sites further into the loop (such as S346C) are closer to each other (37 Å) than sites closer to S4 (V363C at 45 Å). This suggests that the S3-S4 loop may be tilted with respect to the channel's central axis in the closed state. In the open state, distances at these residues suggest that the S3-S4 linker is less tilted (Table 1). If S4 is aligned with the S3-S4 loop, S4 may also undergo a tilt in response to voltage. This hypothesis is also consistent with the larger movements seen at S346C than at sites closer to S4.

Voltage-dependent energy transfer changes were also detected at successive sites near S4, at S351C, S352C and N353C. Unexpectedly, the changes are different for each site (Fig. 4a). If the energy-transfer changes are caused by a physical translation, fluorophores at site S351C move ~ 1 Å further apart as the channel opens, site S352C shows no significant change in distance, and site N353C moves ~ 1 Å closer together. A simple model consistent with this data involves a rotation of this protein region. Fig. 4b portrays this region of the protein as an α -helix, which undergoes a rotation moving S351C further from the pore, N353C closer to the pore, and maintaining S352C at the same distance. It is also possible that the energy-transfer changes at these sites are caused by a reorientation of the acceptor without a corresponding translation. However, either interpretation supports a rotation of these three residues.

Movement in the plane of the membrane, which causes distance changes between subunits, can be measured with our technique. In addition, transmembrane movements can also be measured. At small depolarizations, some voltage sensors are in the activated position, while others are in the deactivated position, giving rise to distance changes between subunits with voltage (Fig. 5a, Active). In contrast, at very hyperpolarized (Fig. 5a, Closed) or depolarized (Fig. 5a, Open) potentials, the distance between subunits would be the same in either state. Assuming that the movements of the voltage sensors are independent with a transmembrane movement of 16 Å, as has been proposed^{7,17}, the intersubunit distance versus voltage

would demonstrate a bell-shaped voltage dependence with a peak change in distance of 2.2 Å at intermediate potentials (Fig. 5b). As the actual voltage-dependent movement has a very different shape (Fig. 3b) and no such bell-shaped voltage-dependent movement was measured in or near S4, it is unlikely that this region undergoes a large non-cooperative transmembrane movement with voltage.

If the S3–S4 region is undergoing a rotation and possible tilt, what is the motion of S4, where the critical charged residues affected by the external electric field reside? S4 could undergo a large translation which, via a kink, leads to a rotation and tilting of S3–S4. However, the lack of distance changes at V363C, which is in S4, argues against this. Hence, we think it more likely that S4 itself is undergoing a rotation which drives the rotation seen in S3–S4.

These results are the first to measure actual voltage-induced distance changes in an ion channel. The movements at the particular residues measured correlate very well with electrical movement of charged residues in the voltage sensor. Our results imply that the voltage sensor need not undergo a large conformational change, as previously proposed^{7,17}. One physical model that is consistent with our data and previous studies is shown in Fig. 5c. In the closed and open states of the channel, residues in and near the S4 segment reside in crevices within the protein for which histidine-associated protons have deeper accessibility than cysteine-reactive reagents^{7,8}. The two crevices must be very close, separated by a non-conducting (hydrophobic) region, formed by the S4 itself. A small

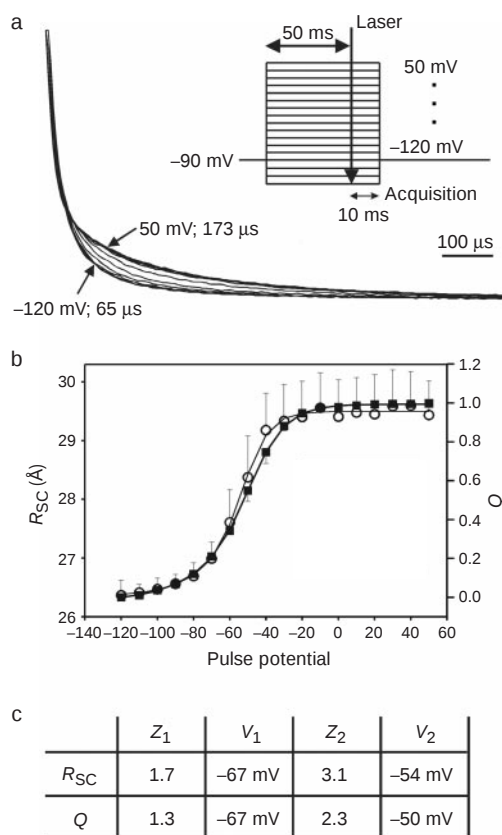


Figure 3 Voltage-dependent distance changes at residue S346C and correlation to gating charge movement. **a**, Voltage-dependent sensitized emission changes measured at site S346C from 65 μ s at -120 mV to 173 μ s at 50 mV, measured 50 ms after a pulse to voltages -120 mV and 50 mV in increments of 10 mV. Inset, protocol. **b**, Movement of S346C between contiguous subunits (R_{SC} , open circles) and gating charge displacement (Q , black squares) as a function of voltage (error bars: standard error, $n = 5$). Both R_{SC} (thin line) and Q (thick line) were fitted to a sequential three-state model (see Methods). **c**, Parameters of the sequential two-Boltzmann fit for R_{SC} and Q , where Z_i and V_i represent the valence and voltage midpoint of the i th Boltzmann (see Methods).

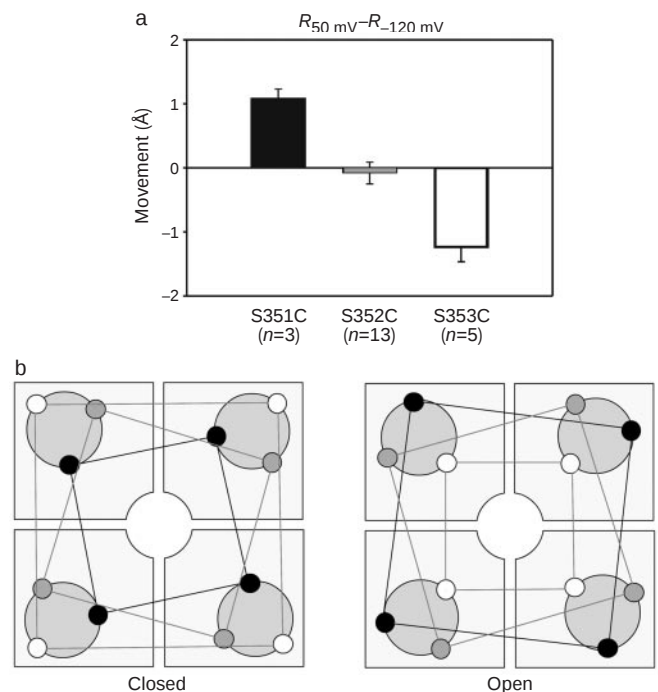


Figure 4 Distance changes at three successive residues in the S3–S4 linker. **a**, Distance changes in R_{SC} measured at successive residues; S351C (black, 1.1 Å further apart at 50 mV), S352C (grey, no change), and N353C (white, 1.3 Å closer together at 50 mV). Changes in distance from the open state ($R_{50\text{ mV}}$) to the closed state ($R_{-120\text{ mV}}$) were determined by measuring sensitized-emission lifetimes at different voltages in the same oocyte, and results from several oocytes were averaged with standard error (bars) shown. **b**, A physical model which leads to the distance modulation seen in **a**. If S351 (black), S352 (grey), and N353 (white) are members of an α -helix which rotates 180° as the channel opens, then S351 will move further apart, S352 will maintain the same distance, and N353 will move closer together. For clarity, the rotation is shown as 180°; however, the actual angle may be considerably less.

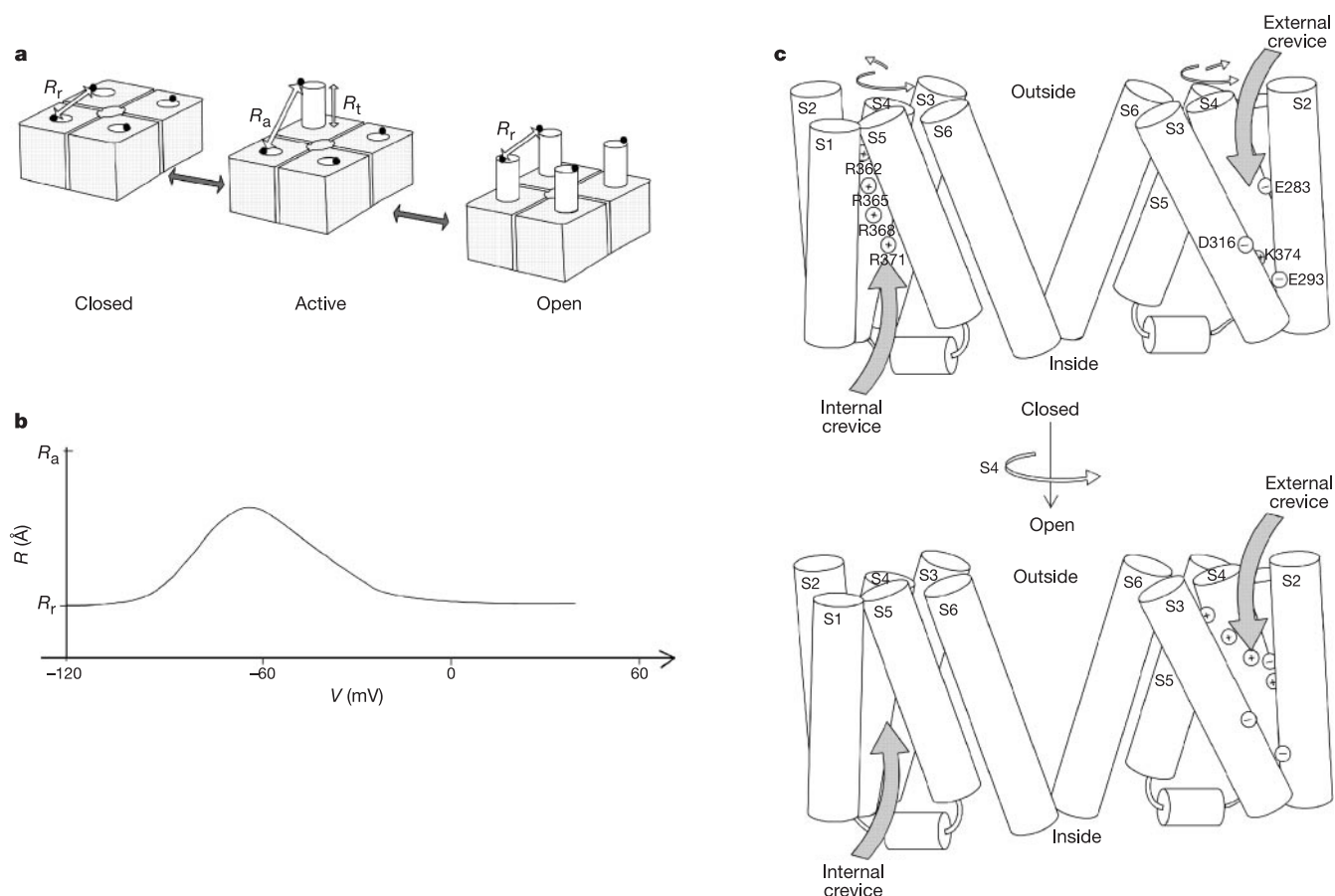


Figure 5 Translation versus rotation models of S4 movement. **a**, Diagram of a large S4 movement perpendicular to the membrane. Each subunit (square block) contains a moveable S4 (cylinder). For the closed or open states (labelled), the distance between contiguous subunits (R_r) will remain constant, but will change to R_a at intermediate potentials (Active), where only one or two S4s are activated. **b**, A large transmembrane movement, as described in **a**, will generate a peak distance change at intermediate potentials which depends on R_r , R_t (distance moved by the S4) and the independence of subunit movement. Assuming independence of voltage sensor movement, $R_r = 30 \text{ \AA}$,

conformational change such as a rotation can then move the charge across the field, while changing chemical accessibility from the intracellular to the extracellular solution.

The present measurements give new insight into the physical rearrangements underlying gating in ion channels. The technique of LRET, particularly when combined with electrical measurements, can be used to study atomic scale structural changes in this channel and, in the future, other transmembrane proteins *in situ*. □

Methods

Distance computation

Energy transfer (E) is calculated by comparing the donor-only lifetime (τ_D) with the lifetime of the sensitized emission (τ_{AD}) using the relation $E = 1 - \tau_{AD}/\tau_D$. The distance (R) is then computed using the conventional formula, $R = R_0(E^{-1} - 1)^{1/6}$ where R_0 is a characteristic distance depending on choice of donor and acceptor: here $R_0 = 45 \text{ \AA}$ (refs 18, 19). Multiple τ_{AD} lifetimes indicate that there are several different donor-acceptor pairs undergoing different amounts of energy transfer from which separate distances can be calculated. For distance changes, the acceptor-sensitized emission was fitted to distances constrained by the Pythagorean theorem.

Stoichiometry of labelling

Given the labelling ratio of 4:1 of terbium maleimide to fluorescein maleimide and a binomial distribution of subunit labelling, 41% of channels will have four donors, 41% will have three donors and one acceptor, 15% will have two donors and two acceptors, 2.6% will have three acceptors and one donor, and 1.6% will have four acceptors. The channels with four donors or four acceptors will not contribute to acceptor-sensitized

emission, as there are no donor-acceptor pairs in these channels. The small subpopulation of channels with two donors and two acceptors will yield different distances, as each donor sees two nearby acceptors instead of one. However, the error is very small for this 15% subpopulation of channels: $\sim 1.4 \text{ \AA}$ for a distance of $28 \text{ \AA}$. This error is also nearly eliminated when measuring differences in distances, such as changes in distance as a function of voltage.

Sources of error in distance measurements

R_0 is affected by the anisotropy of donor and acceptor. The donor is inherently unpolarized (G. E. Snyder & P. R. Selvin, unpublished data). The anisotropy of fluorescein, which is a measure of its nanosecond-scale wobble, was determined at different channel sites, and maximum error estimates resulting from these anisotropy values have been included in the table for each site (see Supplementary Information). Additional wobble during the micro-to-millisecond donor lifetime further reduces this error. The other source of uncertainty is the variability of the fluorophore's position with respect to the labelled residue, owing to the length of the maleimide linker. If the probes either sweep out a large volume, or are far from their attachment sites with the donor sitting at a significantly different position from the acceptor, then the four-fold symmetry breaks down. Consequently, the Pythagorean relationship between distances across (R_{SA}) and contiguous (R_{SC}) to the pore would no longer be valid. In fact, the deviation from the Pythagorean relationship depends on the size of the volume sampled and the flexibility of the linker. For instance, if the linkers permit the fluorophores to sample a volume with a $10 \text{ \AA}$ radius, and they are attached to residues located $50 \text{ \AA}$ apart, then the measured contiguous distance would be $\sim 30 \text{ \AA}$ because the R^{-6} power dependence places the heaviest weight on the closest distance between fluorophores. In that case, the difference between the predicted distance (from Pythagorean theorem) and the measured distance is $7 \text{ \AA}$. However, our measured and predicted distances from the Pythagorean theorem are typically within $1\text{--}2 \text{ \AA}$. The error introduced by flexibility of the linker is therefore $< 2 \text{ \AA}$ and would not account for our measured distance changes.

Experimental protocols

The fast inactivation-removed, non-conducting version of the *Shaker* K⁺ channel (H4IR) was used, and site-directed mutagenesis, transcription of cRNA, and experimental solutions were carried out as described²⁰. Oocytes were rinsed in a sterile SOS without DTT and labelled with 2 μM TbM and 0.5 μM FM for 30 min at 18 °C. Gating, ionic, and fluorescence currents were acquired as described²⁰. The optical setup consisted of a Zeiss Axiovert 10 with a 337-nm nitrogen laser (Laser Science; Newton, MA) as light source, and a 350-nm dichroic with a long pass filter 515 nm in length for donor-only fluorescence or a 510–530 nm filter for sensitized emission (Omega Optical). A custom oocyte chamber was designed to view the clamped oocyte surface in an inverted cut-open configuration using a fused quartz 20× objective with numerical aperture 0.65 (Partec). A voltage electrode was inserted from above into the oocyte interior to measure the transmembrane potential for the virtual ground voltage clamp. Light measurements were made with a gallium arsenide R943-02 photomultiplier tube (PMT) (Hamamatsu), with gate normally on option, operated at 800–1,400 V with an S502 power supply and cooled to –25 °C with a TE104RF PMT cooler (Products for Research). A TTL pulse was used to turn off the PMT during the laser pulse to avoid saturation. The current output of the assembly was amplified with a current-to-voltage converter with a 1-MΩ feedback resistor and was filtered at 20 kHz.

Three-state model

The equation used to fit the distance (R_{SC}) and gating charge displacement (Q) was

$$f(V) = a + \frac{bz_0[1 + \exp(-z_1e(V - V_1)/kT)]}{1 + \exp(-z_1e(V - V_1)/kT)[1 + \exp(-z_0e(V - V_0)/kT)]} + \frac{bz_1}{1 + \exp(-z_1e(V - V_1)/kT)[1 + \exp(-z_1e(V - V_0)/kT)]}$$

where z_0 and z_1 are the valences and V_0 and V_1 are the midpoints of the two sequential steps. For R_{SC} , a is the minimum distance and $b(z_0 + z_1)$ is the maximum distance change. For Q , $a = 0$ and $b(z_0 + z_1) = 1$. This equation reflects the minimum model that is appropriate to fit the charge movement of the voltage sensor²¹.

Model assumptions

It is conceivable, but unlikely, that the voltage sensors in each channel are so highly cooperative that all voltage sensors move simultaneously at all potentials, preventing the intermediate state in Fig. 4b from occurring. Cooperativity in *Shaker* has been shown, but occurs in transitions close to the open state at voltages where most of the gating charge has already moved^{22,23}. Also, most kinetic models of the *Shaker* potassium channel utilize four subunits which behave independently at hyperpolarized potentials, and then undergo a cooperative, concerted transition into the open state²³. Evidence in sodium channels also indicates that at intermediate potentials, some subunits' voltage sensors are activated, while other subunits' voltage sensors are deactivated (see Supplementary Information).

Received 22 July; accepted 19 October 1999.

- Noda, M. *et al.* Primary structure of the *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature* **312**, 121–127 (1984).
- Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, Y. L. & Jan, L. Y. Sequence of a probable potassium channel component encoded at *Shaker* locus of *Drosophila*. *Science* **237**, 770–775 (1987).
- Armstrong, C. M. & Bezanilla, F. Currents related to the movement of the gating particles of sodium channels. *Nature* **242**, 459–461 (1973).
- Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the *Shaker* K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
- Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the *Shaker* K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
- Yang, N. & Horn, R. Evidence for voltage-dependent S4 movement in sodium channels. *Neuron* **15**, 213–218 (1995).
- Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the *Shaker* K⁺ channel S4. *Neuron* **16**, 387–397 (1996).
- Starace, D. M., Stefani, E. & Bezanilla, F. Voltage-dependent proton transport by the voltage sensor of the *Shaker* K⁺ channel. *Neuron* **19**, 1319–1327 (1997).
- Mannuzzu, L. M., Moronne, M. M. & Isacoff, E. Y. Direct physical measure of conformational rearrangement underlying potassium channel gating. *Science* **271**, 213–216 (1996).
- Cha, A. & Bezanilla, F. Characterizing voltage-dependent conformational changes in the *Shaker* K⁺ channel with fluorescence. *Neuron* **19**, 1127–1140 (1997).
- Selvin, P. R. & Hearst, J. E. Luminescence energy transfer using a terbium chelate: improvements on fluorescence energy transfer. *Proc. Natl Acad. Sci. USA* **91**, 10024–10028 (1994).
- Selvin, P. R., Rana, T. M. & Hearst, J. E. Luminescence resonance energy transfer. *J. Am. Chem. Soc.* **116**, 6029–6030 (1994).
- Xiao, M. *et al.* Conformational changes between the active-site and regulatory light chain of myosin as determined by luminescence resonance energy transfer: the effect of nucleotides and actin. *Proc. Natl Acad. Sci.* **95**, 15309–15314 (1998).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Gross, A., Columbus, L., Hideg, K., Altenbach, C. & Hubbell, W. L. Structure of the KcsA potassium channel from *Streptomyces lividans*: A site-directed spin labeling study of the second transmembrane segment. *Biochemistry* **38**, 10324–10335 (1999).
- MacKinnon, R., Cohen, S. L., Kuo, A., Lee, A. & Chait, B. T. Structural conservation in prokaryotic and eukaryotic potassium channels. *Science* **280**, 106–109 (1998).
- Durell, S. R. & Guy, H. R. Atomic scale structure and functional models of voltage-gated potassium channels. *Biophys. J.* **62**, 238–247 (1992).
- Selvin, P. R. Fluorescence resonance energy transfer. *Methods Enzymol.* **246**, 300–344 (1995).

- Chen, J. & Selvin, P. R. Thiol-reactive luminescent chelates of terbium and europium. *Bioconjugate Chem.* **10**, 311–315 (1999).
- Cha, A. & Bezanilla, F. Structural implications of fluorescence quenching in the *Shaker* K⁺ channel. *J. Gen. Phys.* **112**, 391–408 (1998).
- Bezanilla, F., Perozo, E. & Stefani, E. Gating of *Shaker* K⁺ channels: II. The components of gating currents and a model of channel activation. *Biophys. J.* **66**, 1011–1021 (1994).
- Smith-Maxwell, C. J., Ledwell, J. L. & Aldrich, R. W. Uncharged S4 residues and cooperativity in voltage-dependent potassium channel activation. *J. Gen. Phys.* **111**, 421–439 (1998).
- Schoppa, N. E. & Sigworth, F. Activation of *Shaker* potassium channels. III. An activation gating model for wild-type and V2 mutant channels. *J. Gen. Phys.* **111**, 313–342 (1998).
- Selvin, P. R., Jancarik, J., Li, M. & Hung, L. W. Crystal structure and spectroscopic characterization of a luminescent europium chelate. *Inorg. Chem.* **35**, 700–705 (1996).
- Bezanilla, F. The voltage sensor in voltage dependent ion channels. *Physiol. Rev.* (in the press).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by NIH grants to F.B. and P.R.S., the Hagiwara Chair funds to F.B., and from Research Corporation to P.R.S. A.C. was also supported by UCLA Medical Scientist Training Program and a National Research Service Award from National Institute of Mental Health. G.E.S. was supported by a National Research Service Award in Molecular Biophysics. We thank A. Gross, for his involvement in preliminary FRET results, and J. Chen, for synthesis of maleimide terbium chelates. We also thank the members of the Bezanilla lab for their support.

Correspondence and requests for materials should be addressed to F.B. (e-mail: fbezanil@ucla.edu) and P.R.S. (e-mail: selvin@uiuc.edu).

Spectroscopic mapping of voltage sensor movement in the *Shaker* potassium channel

K. S. Glauner, L. M. Mannuzzu, C. S. Gandhi & E. Y. Isacoff

Department of Molecular & Cell Biology, University of California, Berkeley, 271 Life Science Addition, Berkeley, California 94720-3200, USA

Voltage-gated ion channels underlie the generation of action potentials and trigger neurosecretion and muscle contraction. These channels consist of an inner pore-forming domain, which contains the ion permeation pathway and elements of its gates, together with four voltage-sensing domains, which regulate the gates^{1–6}. To understand the mechanism of voltage sensing it is necessary to define the structure and motion of the S4 segment, the portion of each voltage-sensing domain that moves charged residues across the membrane in response to voltage change^{7–14}. We have addressed this problem by using fluorescence resonance energy transfer as a spectroscopic ruler^{15–17} to determine distances between S4s in the *Shaker* K⁺ channel in different gating states. Here we provide evidence consistent with S4 being a tilted helix that twists during activation. We propose that helical twist contributes to the movement of charged side chains across the membrane electric field and that it is involved in coupling voltage sensing to gating.

Opening of voltage-gated channels is triggered by the depolarization-driven outward movement of the S4 segment through a transmembrane canal^{17–14}. This motion generates a small gating current that reflects the movement of charged residues across the membrane electric field. Different types of transmembrane movement have been proposed, leading to distinct models describing how the membrane electric field drops across S4, the forces that S4 needs to overcome to move through the canal and the mechanism of coupling between S4 and the pore domain's gates.

In an attempt to understand how S4 moves we used fluorescence resonance energy transfer (FRET) to measure intersubunit distances between positions within and just outside S4 in both resting and

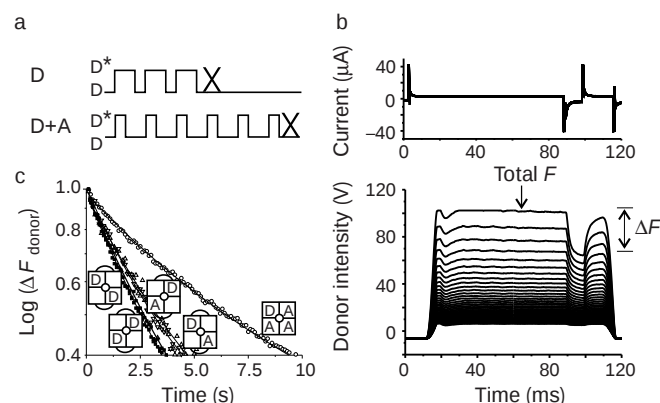


Figure 1 FRET efficiency by indirect donor lifetime measurement. **a**, Photodestruction rate depends on donor (D) cumulative excited lifetime. Donor excited state (D^*) lifetime is shortened by FRET to an acceptor (A), allowing D to undergo more excitation cycles before photodestruction (X), yielding slower macroscopic photobleaching. **b**, Representative photobleaching experiment with gating current (top) and corresponding fluorescence traces (bottom) from a voltage-clamped oocyte expressing *Shaker* S351C labelled with donor FM. Every tenth trace is shown of a total of 411. Absence of change in gating current indicates stability of recording. **c**, Photobleaching of the normalized ΔF_{donor} for channels formed from wild-type S352C or wild-type S352C-linked dimers labelled with D alone (squares) or with D + A in a 1:4 ratio (triangles), and for S352C monomers labelled with D + A in a 1:4 ratio (circles).

activated channels. Distances between labelled sites were determined from the degree to which the excited state lifetime of a donor fluorophore was shortened by energy transfer to an acceptor fluorophore (Fig. 1a), thus slowing the rate of irreversible donor photodestruction (Fig. 1b)¹⁸. Determinations were made exclusively from voltage-clamped channels by measuring the voltage-driven change in fluorescence of the donor fluorophore (ΔF_{donor}) (Fig. 1b, see Methods).

We site-specifically attached the donor (fluorescein maleimide (FM)) and acceptor (tetramethylrhodamine maleimide (TMRM)) fluorophores to single cysteines introduced at an identical site on each subunit¹⁰. Initially, we compared the rate of donor photodestruction ($\Delta F_{donor-bleach}$) at position 352 (in S3–S4) at positive (activating) voltage between donor alone and a 1:4 donor:acceptor mixture. We examined homotetrameric channels that could be labelled on all four subunits, and channels made of linked dimers that could only be labelled at diagonal positions¹⁹ (Fig. 1c). The rate of $\Delta F_{donor-bleach}$ for channels labelled with donor alone was fast and almost identical for the channels made from dimers (Fig. 1c) and the homotetramer (data not shown). The $\Delta F_{donor-bleach}$ slowed modestly in the channels made from dimers, and slowed even more in the homotetramers, in which most of the D-to-A energy transfer is between adjacent subunits (Table 1).

From the measured total FRET efficiency in the homotetramers we calculated the adjacent and diagonal transfer components, and these values were translated to distance using the Foerster equation (see Methods). The fact that donor fluorescence changes with activation at each of our sites suggests that the quantum yield of the donor FM (Φ_D) may differ between gating conformations. Φ_D may also differ between sites, resulting in site and state-dependent differences in the distance corresponding to 50% efficiency for a specific donor–acceptor pair (R_0). We therefore determined Φ_D for each site in the resting and activated states (see Methods), and then calculated values of R_0 using these values of Φ_D (Table 2). Diagonal FRET efficiency calculated from the 352 homotetramers, assuming fourfold symmetry, closely agrees with the measured FRET efficiency in the channels made from dimers (Table 1), where only diagonal interactions are expected.

We calculated distances between adjacent subunits at positions 350, 351, 352, 353 and 356 in the resting state (Table 2, see

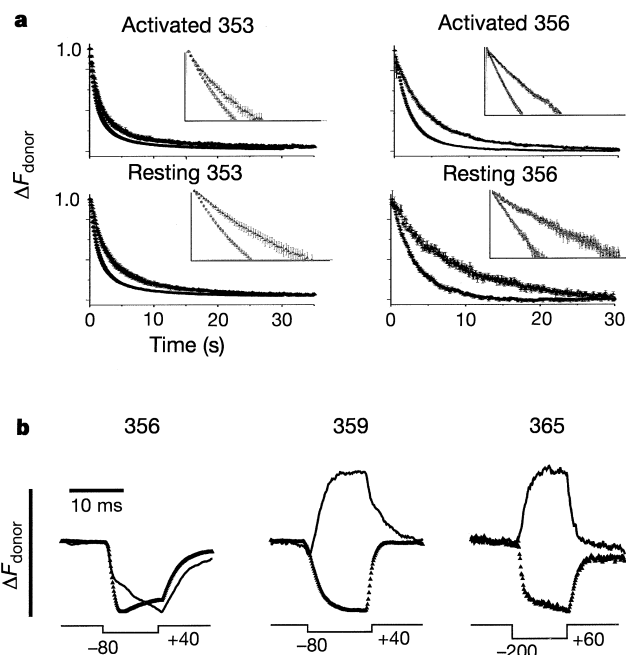


Figure 2 Steady-state and dynamic measurement of FRET efficiency change with activation. **a**, Photobleaching of the normalized ΔF_{donor} at sites 353 and 356. In the presence of an acceptor (D + A), the donor bleaches slower than with donor alone (D), indicating transfer of the donor excited state to the acceptor. Efficiency is higher in the resting state at both 353 and 356. (Mean results for five oocytes of each. Error bars are s.e.m. Inserts are on a semilog scale.) **b**, ΔF_{donor} elicited by hyperpolarizing steps changes direction in presence of the acceptor A (D + A) compared with donor alone (D) for positions 356, 359 and 365. Amplitudes are normalized, with the D + A traces amplified 12× (359) and 7.5× (365).

Methods). Assuming fourfold symmetry, this yielded distances to the central axis of 36.5–43.4 Å, which is consistent with the voltage-sensing domain wrapping around the outer rim of the pore domain, which in KcsA has a radius of about 25 Å at its widest point². FRET efficiency differed between the resting and activated states (Table 2, see positions 353 and 356 in Fig. 2a). With the exception of 351, these sites increase distance upon activation (Table 2).

In addition to quantitative distance information, we also obtained a qualitative indication about distance change from the shape of the ΔF_{donor} trace. At three positions, 356, 359 and 365, the ΔF_{donor} trace changed sign in the presence of acceptor (Fig. 2b). For example, in channels labelled at position 356 with donor alone, deactivating voltage steps gave rise to a two-component ΔF_{donor} consisting of an early phase of dimming followed by a slower and smaller phase that increased in brightness (Fig. 2b). When labelled with both donor and acceptor, the slower ΔF_{donor} component reversed direction and became dimmer, suggesting that quenching by FRET of the resting state relative to the activated state overcomes the increase in brightness due to the change in microenvironment. This agrees with the steady-state determination of higher FRET efficiency in resting channels (Fig. 2a). At positions 359 and 365, a characteristic downward ΔF_{donor} in response to a deactivating voltage step was reversed in the presence of acceptor (Fig. 2b). This shows that positions 359 and 365 are ‘de-quenched’ at rest, displaying an opposite behaviour to position 356. Given the similarity between the resting and activated R_0 values for each of these sites (Table 2), these changes in FRET indicate that donor and acceptor attached at 359 and 365 become closer in the activated state. (The small size of ΔF_{donor} in the presence of the acceptor prevented us from quantifying FRET with photobleaching at positions 359 and 365.)

The real-time measurements of quenching and de-quenching at

Table 1 Donor:acceptor stoichiometry and FRET measurement for position 352 in the activated state

FRET efficiency	D–A interaction	Labelling pattern	No. of donors	Channels (%)	(1–FRET efficiency)	Total fluorescence	Fluorescence (%)
0.0			4	0.2	1.00	0.8	2.3
0.08	(diagonal)		1	0.9	0.92	0.8	2.3
0.42	(adjacent)		2	1.7	0.58	2.0	5.7
0.45	(diagonal + adjacent)		2	10.3	0.55	11.3	32.2
0.59	(adjacent + adjacent)		2	5.1	0.51	4.2	12.0
0.61	(diagonal + adjacent + adjacent)		1	41.0	0.49	16.0	45.6
NA			0	41.0	NA	0.0	0.0

A 1:4 donor:acceptor ratio results in a binomial distribution of labelling stoichiometries. The contribution of donor fluorescence for each stoichiometry is a product of the number of donors, the percentage of channels with that stoichiometry, and the relative brightness of the donor after quenching by FRET to the acceptor(s) (that is, $1 - \text{FRET efficiency}$). Because of the sixth power dependence on distance, total FRET efficiency is dominated by energy transfer between adjacent subunits. Between 90% and 95% of the fluorescence signal is predicted to come from channels with 0, 1 or 2 donors and at least one neighbouring donor–acceptor interaction.

positions 356, 359 and 365 show that the change in donor–acceptor distance takes place in milliseconds. De-quenching with steps to negative voltages at positions 359 and 365 paralleled the kinetics of the gating current (data not shown), indicating that the same motion both moves basic residues on S4 across the membrane electric field and changes intersubunit distances between fluorophores attached to S4.

Photobleaching and kinetic data show that as one advances along S3–S4 and into S4 there is an alternation of sites that are driven closer and farther apart during channel activation (Table 2). This pattern cannot be explained by an outward activation motion of perpendicular S4s, which would keep the mean distance constant, or by a translation along a tilted axis, which would change all of the distances in the same direction. However, the pattern is consistent with an α -helix that twists 180° during activation (Fig. 3a), thus moving four residues on one face (positions 350, 352, 353 and 356) apart with activation and bringing three residues on the opposite face (positions 351, 359 and 365) closer together. In keeping with this idea, the largest movement apart was at 353 and 356, directly across the helix from position 351, which made the biggest motion in the opposite direction. The coherence of the pattern supports the structural interpretation of a helical twist, despite uncertainty in individual distance determinations that arise from fluorophore size (which places the donor and acceptor dipoles at some distance from the protein backbone) and anisotropy (see Methods). The inter-

pretation is consistent with earlier evidence that S4 peptides form an α -helix²⁰.

If an S4 helix were perpendicular to the membrane, then a twist would switch the face of the helix farthest from the pore of the channel with the closest face (for example, positions 353 and 356 would be closest at rest, and 351 closest when activated). Contrary to this simple expectation, the distance from central axis increased linearly from positions 350 to 356, in both resting and activated channels (Fig. 3b), a pattern that suggests that S3–S4 is tilted with its amino-terminal end pointing in the general direction of the central axis.

In what way may a twisting motion of S4 be associated with its outward displacement of the gating charge? If the canal is rigid, then a helical screw-like motion could permit the basic residues found at every third position on S4 to pair with acidic residues and ratchet between them, as previously proposed^{21,22}. As the pitch of the screw would correspond to a spiral stripe of basic residues along the helix, a 180° twist would involve an axial translation of about nine residues, or 13.5 Å (Fig. 3c, top).

Alternatively, as the four main charge-carrying residues (R362, R365, R368 and R371)^{7,9,11–14} lie on one face of a helix (Fig. 3a), a 180° twist might on its own transfer this charged face from the internal to the external vestibule of the canal. Such a motion could account for much of the gating charge displacement, decreasing the need for a large axial translation. The smaller the axial translation,

Table 2 Summary of measurement and analysis for all sites

Site	Anisotropy (acceptor)		Φ_D		R_0 (Å)		Transfer efficiency					
							Total measured (%)		Adjacent calculated (%)		Distance to adjacent (Å)	
	Rest	Act.	Rest	Act.	Rest	Act.	Rest	Act.	Rest	Act.	Rest	Act.
350	0.212 ± 0.010 (7)	0.207 ± 0.008 (7)	0.30 ± 0.01 (4)	0.22 ± 0.01 (4)	50.4 ± 0.3	48.1 ± 0.2	60.0 ± 2.8 (15)	53.3 ± 1.9 (23)	47.0	38.0	51.6	52.2
351	0.182 ± 0.001 (7)	0.164 ± 0.001 (7)	0.59 ± 0.02 (4)	0.36 ± 0.02 (4)	57.7 ± 0.3	52.0 ± 0.5	55.7 ± 1.5 (19)	56.2 ± 1.9 (23)	41.0	42.0	61.3	54.9
352	0.109 ± 0.014 (7)	0.100 ± 0.012 (7)	0.45 ± 0.02 (5)	0.36 ± 0.01 (5)	54.1 ± 0.3	52.0 ± 0.2	65.9 ± 1.3 (23)	56.5 ± 1.6 (22)	52.0	42.0	53.4	54.9
353	0.149 ± 0.005 (7)	0.133 ± 0.005 (7)	0.42 ± 0.03 (4)	0.47 ± 0.02 (4)	53.4 ± 0.7	54.5 ± 0.4	60.8 ± 1.3 (14)	53.3 ± 1.9 (15)	47.5	38.0	54.3	59.1
356	0.190 ± 0.004 (6)	0.186 ± 0.004 (6)	0.56 ± 0.01 (5)	0.55 ± 0.03 (4)	56.0 ± 0.2	55.9 ± 0.5	60.7 ± 1.9 (20)	51.1 ± 1.5 (23)	47.5	35.6	57.0	61.7
359	0.102 ± 0.023 (7)	0.161 ± 0.007 (7)	0.44 ± 0.05 (4)	0.58 ± 0.05 (4)	53.8 ± 1.1	56.4 ± 0.9	–	–	–	–	–	–
365	0.154 ± 0.003 (7)	0.162 ± 0.002 (7)	0.22 ± 0.01 (4)	0.16 ± 0.01 (4)	48.1 ± 0.4	45.5 ± 0.5	–	–	–	–	–	–

Acceptor anisotropy, donor quantum yield (Φ_D), R_0 , FRET efficiency and estimated distance were determined as described in Methods. Changes with activation in distance between adjacent subunits were 3–13-fold larger than the error in R_0 that resulted from experimental scatter in the determinations of Φ_D . Changes in distance followed changes in FRET efficiency, except at position 351 where the resting and activated distances differed despite similar FRET efficiency because of a very significant difference in Φ_D ($P < 0.00003$). Values are mean ± s.e.m. Number of oocytes is given in parentheses.

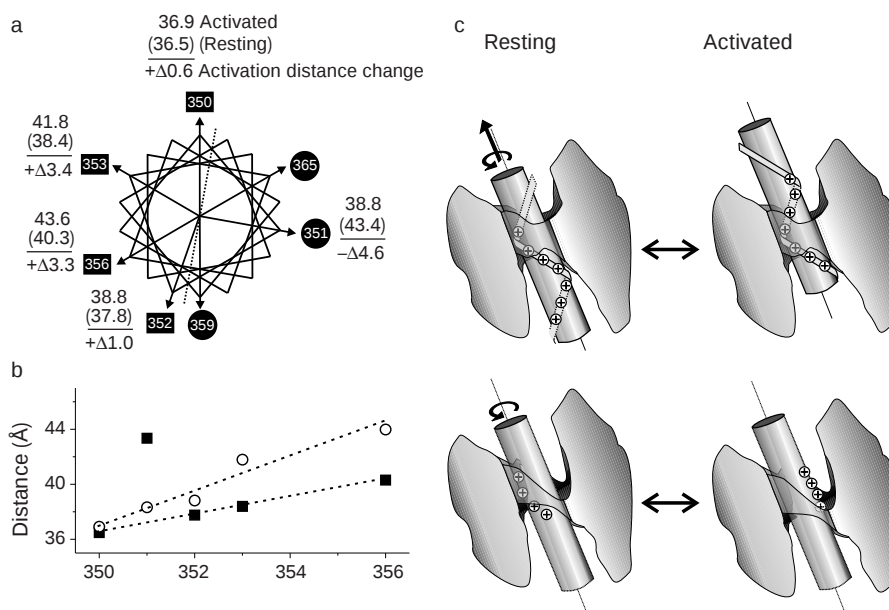


Figure 3 Activation motion of S4 can be accounted for by a helical twist of 180°. **a**, Distances calculated from FRET efficiencies projected onto a helical wheel. Distances to central axis calculated for resting (in parenthesis) and activated states (with change in distance (Δ)) assuming fourfold symmetry and using the Pythagorean theorem. Activation moves rectangular sites apart and circled sites closer. **b**, Distance to central axis versus

the deeper the canal vestibules would need to penetrate on the opposite faces of S4 (Fig. 3c, bottom). To enable basic residues of S4 that are in the canal to pair always with acidic residues from the canal wall, an S4 twist without the full axial translation along the spiral stripe would demand a change in canal conformation. Because S5 and S6 probably contribute to one wall of the canal, such a rearrangement could transmit force to the pore gates.

Both models are compatible with a short canal that permits the cysteine side chain and the fluorophore linker to extend the fluorophore far enough out of the canal to provide reasonable rotational freedom, as seen from the measures of anisotropy in both the resting and activated states (Table 2). Both models are also consistent with previous studies that probed S4 exposure at positions that are mostly in register with the stripe of basic residues^{8,9,11–14}.

A 180° twist would be expected to place substantial torque on the S3–S4 and S4–S5 linkers, especially on the short and conserved S4–S5 (ref. 23), which could transmit torque to S5. This is a significant point, as a twist of the KcsA analogues of S5 and S6 appears to open that channel⁵ and the internal gate of *Shaker* may operate in a similar way⁶. At the same time, a twist of S4–S5 could rearrange the internal mouth of the pore to form the docking site for the inactivation ball²⁴. These pore domain motions could occur cooperatively in all of the subunits under the combined strain generated by the four S4s. In summary, our results are consistent with S4 being a helix that undergoes a 180° twist during its outward charge-carrying movement. A helical secondary structure of S4 would have the effect of reducing the length of the canal, thus focusing the membrane electric field on S4. We propose that transmission of S4 twist to the pore domain may couple voltage sensor motion to opening of the internal mouth of the pore. □

Methods

Molecular Biology

Experiments were performed on non-conducting, ball-deleted ShH4 (W434F/Δ6–46) *Shaker* channels after removal of two native cysteines (C245V and C462A)¹⁰. Site-directed mutagenesis, dimer fusion gene construction, complementary RNA synthesis and cRNA injection into *Xenopus* oocytes were done as described^{10,14,19}.

position in resting (squares) and activated (circles) states. **c**, Top, helical screw model of S4. A 180° twist yields an axial translation of nine residues. Bottom, model of S4 twist without axial translation. A 180° twist with no outward motion moves four charged residues, which span 180° of an α -helix, part way or the entire way from the boundary of the internal vestibule to the external vestibule.

Voltage-clamp FRET

Voltage-clamp fluorometry and oocyte preparation and incubation were done as described^{10,14}, except that oocytes were labelled by incubation with either 25 μ M FM, 100 μ M TMRM, or the mixture 25 μ M FM + 100 μ M TMRM, for 30 min on ice. These concentrations are in excess of the concentration found to saturate labelling. A donor:acceptor ratio of 1:4 favoured a labelling stoichiometry with multiple acceptor subunits and zero or one donor subunit (Table 1). The optical filter combination (excitation 475/40, D495, emission 522/40) made the magnitude of the acceptor signal less than 0.005, and the greater photolability of fluorescein compared with rhodamine ensured that negligible bleaching of rhodamine occurred, providing a pure measure of donor photobleaching. Oocytes were voltage clamped at either positive (40 mV) or negative (–80 mV) voltage and illuminated for 100 ms once they had completed activating or deactivating. The test voltage remained for 90% of the illumination time, with the interruption of a 10-ms step to evoke the ΔF_{donor} . Episodes were separated by 900 ms to allow recovery from inactivation. The difference between FRET measurements in the resting and activated protocols was an underestimate because the protocols used had a 90% duty cycle and the resting protocol at –80 mV does not fully retract the gating charge¹⁴. Determinations were made exclusively from channels controlled by the voltage clamp by measuring ΔF_{donor} in response to a voltage step (Fig. 1b). In this way fluorescence from endocytosed labelled channels and labels on other proteins, and autofluorescence were eliminated from the measurements.

FRET measurement by photobleaching

Distances between labelled sites were determined from the degree to which the excited state lifetime of a donor fluorophore was shortened by energy transfer to an acceptor fluorophore. Donor lifetime was measured indirectly from the rate of irreversible donor photodestruction¹⁸. Traces were baseline adjusted using the fluorescence just before the voltage step. The fluorescence at the end of the step was then plotted versus time (Fig. 2a). Data points were fit to an exponential function to estimate the fluorophore bleach rate. For most cases, data were well described by a single exponential. If necessary, data were fit to a double exponential, and the time-constant corresponding to the fast component was used. FRET efficiency (E) was calculated from the ratio of the donor bleach rates in the absence and presence of the acceptor ($E = 1 - \tau_{\text{DA}}/\tau_{\text{D}}$, where τ_{DA} and τ_{D} are the time-constants of donor photobleaching in the presence and absence of acceptor, respectively). FRET efficiency was converted to distance using Foerster's equation: $E = 1/(1 + R_0^6/R^6)$, where R is the distance between donor and acceptor, and R_0 is the distance corresponding to 50% efficiency for a specific donor–acceptor pair. R_0 is defined by the equation: $R_0 = (9.7 \times 10^3 J \Phi_{\text{D}} n^{-4} \kappa^2)^{1/6}$ (in Å)¹⁷, where J is the normalized spectral overlap of the donor emission and acceptor absorption, Φ_{D} is the donor emission quantum yield in absence of the acceptor, n is the index of refraction and κ^2 is the orientation factor for a dipole–dipole interaction. The values were taken to be $R_0 = 55$ Å for a Φ_{D} of 0.5 (ref. 17); $n = 1.4$; and $\kappa^2 = 2/3$ (ref. 15). As Φ_{D} is directly proportional to the donor excited state lifetime¹⁶, which, in turn, is directly proportional to the donor photodestruction rate, Φ_{D} for each labelled site was calculated from the donor bleach rates (measured at all sites in both activated and resting states in a single experiment), with the assumption that the Φ_{D}

of the fastest bleaching site (351C in resting state) is the same as the Φ_D measured for the donor FM dissolved in recording solution ($\Phi_D = 0.59$, determined using FM in 0.1 N NaOH as a standard¹⁶). The R_0 (mean $\pm$ s.e.m.) for each labelled site was calculated taking into account this estimated Φ_D (mean $\pm$ s.e.m.). Because changes in donor and acceptor fluorescence with gating are not accompanied by shifts in donor emission or acceptor excitation spectra²⁵, J was held constant. Any error in the above assumptions would affect all sites and states in the same way, shifting the values of R_0 in parallel, but not changing the conclusions about differences in distance between sites and states. κ^2 was assumed to be 2/3, as measurements of anisotropy indicated that the attached dyes had reasonable mobility (Table 2). Assuming similar steric constraints on the donor and acceptor dyes, based on their structural similarity, permits estimation of the range of possible values of κ^2 from acceptor anisotropy: 0.42–1.05 for the site with lowest anisotropy, and 0.25–1.60 for the site with the highest anisotropy²⁶. These ranges translate into worst-case scenarios of possible error of 8% and 17% ± 3.2 Å and, corresponding to ± 6.8 Å, respectively, for a distance of 40 Å.

Calculations of FRET efficiency

To translate our measured rates of photobleaching into FRET efficiencies between adjacent and diagonal D–A pairs, we created a 'look-up' table (Table 1) for a range of possible diagonal efficiencies, calculated the corresponding adjacent, and worked out the D–A interactions for each labelling stoichiometry predicted by a binomial distribution for a 1:4 donor:acceptor labelling. An example of this is shown for a diagonal efficiency of 0.08. More than 90% of the donor signal comes from donors with at least one adjacent interaction. The bleach signals from these donors sum to yield a bleach signal that can be well fit by a single exponential with an intermediate time constant that is weighted by the relative amplitudes of the components. The weighted total efficiency is 0.57 for the example in Table 1, closely approximating the total measured percentage of transfer efficiency measured for site 352 in the activated state (Table 2). On the basis of this, the adjacent calculated percentage of transfer efficiency (Table 2) for activated 352 is 42% (Table 1, FRET efficiency), allowing the adjacent distance to be calculated.

Anisotropy measurements

Oocytes injected with individual cysteine mutants and labelled with the acceptor TMRM were voltage clamped in either the resting or activated state, excited with plane polarized light, and the parallel and perpendicular components of the fluorescence measured. Fluorescence anisotropy was calculated using the equation $A = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2I_{\perp})$ (ref. 17), where $I_{\parallel}$ and $I_{\perp}$ are the intensities of parallel and perpendicular fluorescence, respectively.

Received 2 September; accepted 3 November 1999.

- Kubo, Y., Baldwin, T. J., Jan, Y. N. & Jan, L. Y. Primary structure and functional expression of a mouse inward rectifier potassium channel. *Nature* **362**, 127–133 (1993).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- MacKinnon, R., Cohen, S. L., Kuo, A., Lee, A. & Chait, B. T. Structural conservation in prokaryotic and eukaryotic potassium channels. *Science* **280**, 106–109 (1998).
- Perozo, E., Cortes, D. M. & Cuello, L. G. Three-dimensional architecture and gating mechanism of a K^+ channel studied by EPR spectroscopy. *Nature Struct. Biol.* **5**, 459–469 (1998).
- Perozo, E., Cortes, D. M. & Cuello, L. G. Structural rearrangements underlying K^+ channel activation gating. *Science* **285**, 73–78 (1999).
- Yellen, G. The moving parts of voltage-gated ion channels. *Q. Rev. Biophys.* **31**, 239–295 (1998).
- Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K^+ channel. *Neuron* **16**, 1169–1177 (1996).
- Yang, N. & Horn, R. Evidence for voltage-dependent S4 movement in sodium channels. *Neuron* **15**, 213–218 (1995).
- Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the Shaker K^+ channel S4. *Neuron* **16**, 387–397 (1996).
- Mannuzzu, L. M., Moronne, M. M. & Isacoff, E. Y. Direct physical measure of conformational rearrangement underlying potassium channel gating. *Science* **271**, 213–216 (1996).
- Yang, N., George, A. L. Jr & Horn, R. Molecular basis of charge movement in voltage-gated sodium channels. *Neuron* **16**, 113–122 (1996).
- Yusaf, S. P., Wray, D. & Sivaprasadarao, A. Measurement of the movement of the S4 segment during the activation of a voltage-gated potassium channel. *Pflügers Arch.* **433**, 91–97 (1996).
- Starace, D. M., Stefani, E. & Bezanilla, F. Voltage-dependent proton transport by the voltage sensor of the Shaker K^+ channel. *Neuron* **19**, 1319–1327 (1997).
- Baker, O. S., Larsson, H. P., Mannuzzu, L. M. & Isacoff, E. Y. Three transmembrane conformations and sequence-dependent displacement of the S4 domain in Shaker K^+ channel gating. *Neuron* **20**, 1283–1294 (1998).
- Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846 (1978).
- Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry. Part II: Techniques for the Study of Biological Structure and Function* (W. H. Freeman, New York, 1980).
- Van Der Meer, B. W., Coker, G. & Chen, S. Y. D. *Resonance Energy Transfer: Theory and Data* (Wiley, New York, 1994).
- Jovin, T. M. & Arndt-Jovin, D. J. in *Cell Structure and Function by Microspectrofluorometry* (eds Kohen, E., Hirschberg, J. G. & Ploem, J. S.) 99–117 (Academic, New York, 1989).
- Isacoff, E. Y., Jan, Y. N. & Jan, L. Y. Evidence for the formation of heteromultimeric potassium channels in *Xenopus* oocyte. *Nature* **345**, 530–534 (1990).
- Peled-Zehavi, H., Arkin, I. T., Engelman, D. M. & Shai, Y. Coassembly of synthetic segments of Shaker K^+ channel within phospholipid membranes. *Biochemistry* **35**, 6828–6838 (1996).
- Catterall, W. A. Structure and function of voltage-gated ion channels. *Annu. Rev. Biochem.* **64**, 493–531 (1995).

- Durell, S. R., Hao, Y. & Guy, H. R. Structural models of the transmembrane region of voltage-gated and other K^+ channels in open, closed, and inactivated conformations. *J. Struct. Biol.* **121**, 263–284 (1998).
- Fretch, G. C., VanDongen, A. M. J., Schuster, G., Brown, A. M. & Joho, R. H. A novel potassium channel with delayed rectifier properties isolated from rat brain by expression cloning. *Nature* **340**, 642–645 (1989).
- Isacoff, E. Y., Jan, Y. N. & Jan, L. Y. Putative receptor for the cytoplasmic inactivation gate in the Shaker K^+ channel. *Nature* **353**, 86–90 (1991).
- Cha, A. & Bezanilla, F. Characterizing voltage-dependent conformational changes in the Shaker K^+ channel with fluorescence. *Neuron* **19**, 1127–1140 (1997).
- Dale, R. E., Eisinger, J. & Blumberg, W. E. The orientation freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys. J.* **26**, 161–193 (1979).

Acknowledgements

We thank L. Llewellyn for making the linked dimers, and A. Glazer, H. Lecar, J. Ngai, E. Loots, O. Baker, H. P. Larsson, M. Moronne and all the other members of the laboratory for helpful discussions. This work was supported by grants from NIH, American Heart Foundation, CA Affiliate, Department of Energy and Lawrence Berkeley National Laboratory Physical Bioscience Division.

Correspondence and requests for materials should be addressed to E.Y.I. (e-mail: eisacoff@socrates.berkeley.edu).

Functional characterization of a potassium-selective prokaryotic glutamate receptor

Guo-Qiang Chen*, Changhai Cui†, Mark L. Mayer† & Eric Gouaux*

* Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street, New York, New York 10032, USA

† Laboratory of Cellular and Molecular Neurophysiology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

Ion channels are molecular pores that facilitate the passage of ions across cell membranes and participate in a range of biological processes, from excitatory signal transmission in the mammalian nervous system to the modulation of swimming behaviour in the protozoan *Paramecium*¹. Two particularly important families of ion channels are ionotropic glutamate receptors (GluRs)² and potassium channels^{3,4}. GluRs are permeable to Na^+ , K^+ and Ca^{2+} , are gated by glutamate, and have previously been found only in eukaryotes². In contrast, potassium channels are selective for K^+ , are gated by a range of stimuli, and are found in both prokaryotes and eukaryotes^{3,4}. Here we report the discovery and functional characterization of GluR0 from *Synechocystis* PCC 6803, which is the first GluR found in a prokaryote. GluR0 binds glutamate, forms potassium-selective channels and is related in amino-acid sequence to both eukaryotic GluRs and potassium channels. On the basis of amino-acid sequence and functional relationships between GluR0 and eukaryotic GluRs, we propose that a prokaryotic GluR was the precursor to eukaryotic GluRs. GluR0 provides evidence for the missing link between potassium channels and GluRs, and we suggest that their ion channels have a similar architecture, that GluRs are tetramers and that the gating mechanisms of GluRs and potassium channels have some essential features in common.

A search of the non-redundant protein database at the National Center for Biotechnology Information using the Ψ -BLAST algorithm⁵ and the first three membrane-associated segments of the rat GluR2 AMPA receptor² revealed a hypothetical protein from *Synechocystis* PCC 6803 at locus D90909 (ORF_ID:slr1257)⁶. This protein was composed of 397 amino acids, a signal peptide and three hydrophobic segments, flanked by two regions of roughly 140

amino acids (Fig. 1). When the *Synechocystis* protein was used to reprobe the database, the most similar sequence was the rat $\delta 1$ GluR (E-value $\sim 10^{-10}$)^{5,7}, followed by a predicted GluR from *Arabidopsis thaliana*⁸ and numerous eukaryotic GluRs including AMPA receptors^{9,10}. Because of the amino-acid sequence similarities and functional relationships, we have named the *Synechocystis* PCC 6803 protein GluR0.

GluR0 contains regions before and after the transmembrane segments that share significant amino-acid similarity to the S1 and S2 fragments of eukaryotic GluRs (Fig. 1)^{2,11–13}. In eukaryotic GluRs, S1 and S2 form the ligand-binding core^{11,12}. Residues essential to the function and structure of GluR2 and other AMPA receptors^{13,14} are conserved in GluR0: Arg 117 and Asp 314 are positioned such that they can interact with the α -carboxy and α -amino groups of L-glutamate, respectively, and Tyr 86, Pro 110 and Thr 114 flank the putative ligand-binding site; the corresponding residues in the GluR2 S1S2 kainate crystal structure make important interactions with the agonist¹³. Conserved residues in GluR0 and GluR2 also occupy structurally critical regions at the juncture of domains 1 and 2 (residues Phe 125, Pro 128 and Gly 343) and at the end of helix K (Trp 376). The strict conservation of Leu 361, which is located at the site of an exposed hydrophobic patch on helix J of the GluR2 S1S2 structure¹³, suggests that elements of subunit–subunit interactions are conserved between GluR0 and eukaryotic GluRs. However, GluR0 has neither the N-terminal domain (relative molecular mass $\sim 45,000$; $M_r \approx 45K$) present in eukaryotic GluRs nor the third transmembrane segment and accompanying cytoplasmic domain.

When the transmembrane regions of GluR0 (residues Ser 159 to Ala 250) were used to probe the amino-acid sequence database, the most similar sequences were from voltage-dependent potassium channels (E-value, 7×10^{-6})¹⁵ and the KcsA protein from *Streptomyces lividans* (E-value, 3×10^{-5})¹⁶. GluR0 has a canonical potassium-channel selectivity filter sequence (TVGYG) between the two predicted transmembrane segments (Fig. 1). When the selectivity filter sequence was masked and the database search

repeated, the most closely related sequences were eukaryotic GluRs, such as AMPA receptors. Within the P-region and M2 segments, GluR0 has 44% (two gaps) and 41% (no gaps) amino-acid sequence identity to rat GluR2 and KcsA, respectively. In GluR0, the conserved or conservatively substituted residues are localized to key regions of the KcsA structure¹⁷: the pore helix (Trp 200), the selectivity filter (Gly 210) and the end of helix M2 (Ser 238, Thr 240, Ala 241, Leu 243, Ala 244, Phe 247), a portion of the KcsA structure implicated in gating^{17,18}, and a region that produces spontaneous GluR activity and neurodegeneration in Lurcher mice¹⁹ when mutated in $\delta 2$ GluR. Our data support earlier ideas about the relationships between the architecture of the pore regions of GluRs and potassium channels^{20,21}. Nevertheless, there are still important structural and functional differences between eukaryotic GluRs and potassium channels.

In our search for GluR0 ligands, we took advantage of the fact that the GluR0 S1S2 construct was overexpressed by *Escherichia coli* in a water-soluble form. Analogous S1S2 constructs of other GluRs have pharmacological properties that are similar to those of wild-type, membrane-bound receptors^{12,14,22}. Initial ligand-binding studies with ³H-AMPA, ³H-kainate, ³H-(2S,4R)-4-methyl-glutamate or ³H-5,7-dichlorokynurenic acid indicated that none of these classical eukaryotic GluR ligands binds to GluR0. Because the homology between GluR0 and eukaryotic GluRs indicates that amino acids are potential agonists, we next screened a library of ¹⁴C-labelled amino acids in a filter-binding assay (15 in total, not including Cys, Met, Gln, Asn or Trp), using single cold amino acids at a concentration of 1 mM as competitors. In combination with results from further screening experiments involving ¹⁴C-L-glutamine or ¹⁴C-L-tryptophan, these results indicate that L-glutamate has the highest affinity for GluR0. Subsequent ligand-binding experiments using ³H-L-glutamate, individual cold amino acids and cold eukaryotic GluR agonists and antagonists defined a class of compounds that compete for ³H-L-glutamate binding to GluR0 S1S2 (Fig. 2). Although a wide range of compounds compete for ³H-L-glutamate binding to GluR0 S1S2, subtype-selective eukaryotic GluR agonists, such as AMPA,

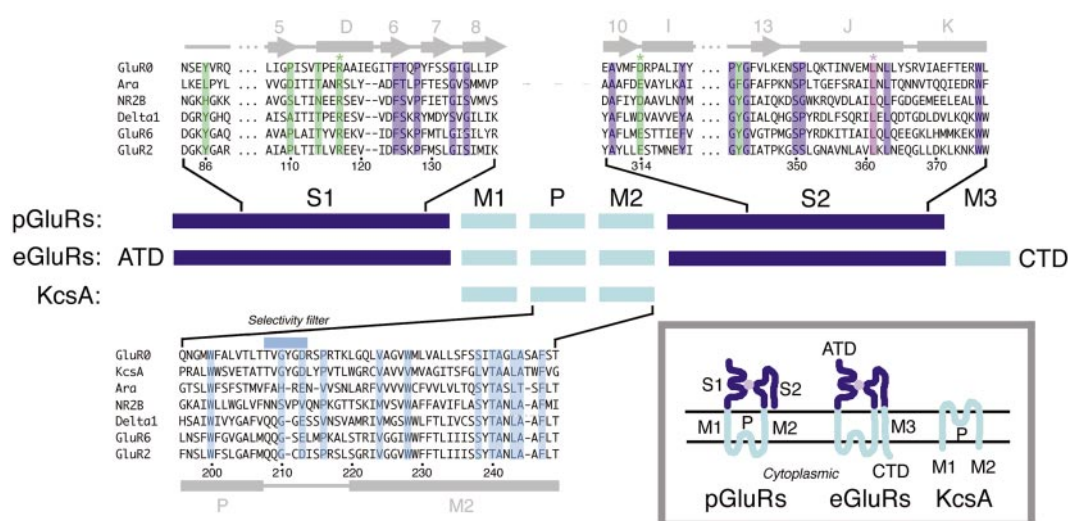


Figure 1 GluR0 is related to eukaryotic GluRs (eGluRs) and to potassium channels. Domain structures for prokaryotic GluRs (pGluRs) such as GluR0 and eukaryotic GluRs and for the transmembrane segments of the KcsA potassium channel are shown in the centre of the Figure. Extracellular portions of GluRs that comprise the ligand-binding core (S1 and S2) are shown in purple and the membrane-associated regions are turquoise. Top, alignments of the sequences of selected GluRs in structurally or functionally important areas. Conserved or conservatively substituted residues are highlighted: residues in or near the ligand-binding site are green; residues in key structural areas are violet; and Leu 361, which is predicted to form an exposed hydrophobic patch on helix J, is pink.

Bottom left, alignments of the P and M2 regions of GluR0, KcsA and eukaryotic GluRs, with conserved or conservatively substituted residues highlighted in blue. A blue bar is drawn over the residues comprising the selectivity filter in GluR0 and KcsA. Shown above S1 and S2 and below P and M2 are elements of secondary structure derived from the GluR2 S1S2 and KcsA structures, respectively^{13,17}. Bottom right, the predicted membrane topology of pGluRs (GluR0) and the membrane topologies of eukaryotic GluRs and KcsA; glutamate is shown as a pink circle. Sequences are defined as follows: GluR0, NCBI *Entrez* accession number 1652933; Ara (a putative eukaryotic GluR from *Arabidopsis thaliana*); NR2B (rat); $\delta 1$ (rat); GluR6 (rat); GluR2 (rat, flop)²; KcsA (*Streptomyces lividans*)³.

kainate and NMDA, do not. For effective interaction with GluR0, amino acids with an α -carbon substituent required L stereochemistry. Variations in the other elements of L-glutamate were more forgiving: even glycine, which is an agonist for the NMDA receptor NR1 subunit, competes for ^3H -L-glutamate binding.

L-Glutamate is the ligand with the highest affinity found so far, and its K_d of ~ 215 nM is similar to the affinity of glutamate for S1S2 constructs of eukaryotic GluRs^{12,22} (Fig. 2c). In contrast to eukaryotic GluRs, L-glutamine effectively competed for L-glutamate binding with a 50% inhibiting concentration (IC_{50}) of 16.3 μM . L-Homocysteic acid (HCA; IC_{50} , 1.6 μM), L-serine (IC_{50} , 30.0 μM) and glycine (IC_{50} , 303 μM) also compete for binding. L- α -Amino adipic acid (L-AA; IC_{50} , 30 μM), which has one more methylene group than does glutamate, competes modestly for ^3H -L-glutamate binding; D, L- α -aminopimelic acid (APA; IC_{50} , 224 μM), which has two more methylene groups, binds roughly 10-fold more weakly still, and L-aspartate binds with even lower affinity. The most important determinants for ligand binding to GluR0 are therefore an L stereochemistry at the α -carbon, and either a carboxylate, a

sulphonate or a primary amide group separated by two methylene units from the α -carbon.

When we had identified a panel of compounds that bind to the receptor, we tried to determine whether L-glutamate, for example, evoked membrane currents in HEK 293 cells and *Xenopus* oocytes assayed using standard voltage-clamp methods. Initial efforts using the wild-type GluR0 coding sequence did not result in ligand-dependent currents. We suspected that poor expression of the prokaryotic GluR0 gene might limit the amount of receptor reaching the plasma membrane, so the first 39 residues of GluR0, which includes the native signal peptide, were replaced by the signal peptide and the 5' untranslated region sequence of rat GluR6, a kainate-preferring eukaryotic GluR that is abundantly expressed in HEK 293 cells and *Xenopus* oocytes.

GluR0 that was modified with the GluR6 sequence produced currents in both HEK 293 cells and oocytes when L-glutamate and other agonists were applied (Figs 3, 4). In HEK 293 cells held at -80 mV, application of 100 μM L-glutamate in a solution containing 50 mM K^+ and 100 mM Na^+ resulted in inward currents that were large enough for functional analysis in $\sim 20\%$ of cells with robust fluorescence from enhanced green fluorescent protein; in 21 cells, the mean response was 101 ± 12.2 pA. In contrast, the response for GluR2 in parallel experiments was 4.0 ± 1.3 nA ($n = 6$). The kinetics of GluR0 responses in HEK cells are relatively slow (Fig. 3): the 10–90% rise time of currents evoked by a 1-s application of 100 μM L-glutamate was 287 ± 14 ms ($n = 7$), whereas for rat GluR2 the value was 18.5 ± 3.0 ms ($n = 7$). Deactivation of GluR0 after 100 μM L-glutamate had been applied also showed slow kinetics (time constant, 191 ± 18 ms; $n = 7$), whereas for GluR2 the current decayed much more rapidly (time constant, 24 ± 3.1 ms; $n = 7$). Longer applications of 100 μM L-glutamate evoked slowly desensitizing GluR0 responses (time constant, 3.5 ± 0.19 s; steady state/peak, 0.29 ± 0.02 ; $n = 13$) (Fig. 3). By comparison, AMPA receptors show much stronger desensitization, with faster and subtype-specific desensitization time constants of between 0.6 and 10.0 ms (ref. 2).

GluR0 does not respond to AMPA, kainate or NMDA, but it is activated by HCA, L-glutamine and, to a lesser extent, by L-serine, L-AA and APA (Fig. 3). In contrast, GluR2 was not activated by L-serine or APA, and was only weakly activated by L-AA. The robust activation of GluR0 by 100 μM L-glutamine ($45.6 \pm 5.2\%$ of the response to 100 μM L-glutamate; $n = 17$) is in contrast to the much weaker response of GluR2 to 100 μM L-glutamine. Dose–response analysis for GluR0 gave a half-maximal effective concentration (EC_{50}) for L-glutamate of 32 ± 2.6 μM ($n = 7$), which was similar to that for GluR2 (21 ± 1.0 μM ; $n = 5$). In the case of GluR0, the discrepancy between the EC_{50} and K_d for L-glutamate (34 μM and 0.22 μM , respectively) may be due to several factors, including the fact that the binding studies probed the affinity of the desensitized state for L-glutamate, rather than the activated form of the receptor. A requirement for glutamate to occupy several GluR0 subunits and a low probability of ion-channel gating would also give EC_{50} values greater than the small K_d for the binding of glutamate to individual subunits.

GluR0 encodes a glutamate-gated channel that exhibits the ion selectivity properties of a potassium channel. Figure 4a shows the maximum current responses at -60 mV of GluR0 and GluR2 expressed in oocytes as a function of the composition of the extracellular solution. In contrast to GluR2, which is relatively non-selective for monovalent cations and insensitive to block by 0.5 mM Ba^{2+} , GluR0 responses were largest in the presence of K^+ and Rb^+ and were abolished when K^+ was replaced by Na^+ or Cs^+ . GluR0 conductance was sensitive to submillimolar concentrations of Ba^{2+} , like that of many other potassium channels⁴. Further support for the conclusion that the GluR0 channel is highly selective for K^+ was obtained from the measurements illustrated in Fig. 4b, showing that the reversal potential of GluR0 varied as a function of the

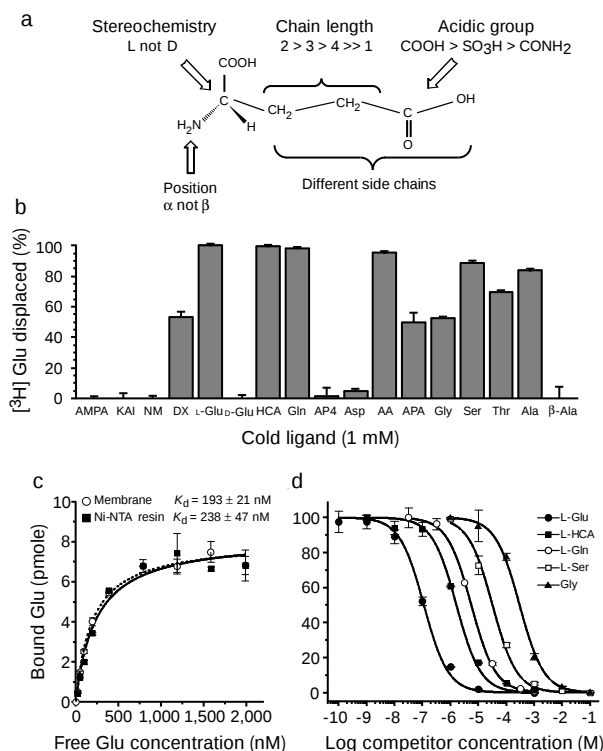


Figure 2 GluR0 binds L-glutamate, L-glutamine and other amino acids. **a**, Different parts of the glutamate structure were varied to probe ligand structure–activity relationships: D-glutamate, α -carbon stereochemistry; β -Ala, location of amino group; L-aspartate, L- α -amino adipic acid (L-AA), and D, L- α -aminopimelic acid (APA), chain length; L-glutamine, L-AP₄ and L-homocysteic acid (HCA), carboxylic acid group; the 20 amino acids, side-chain composition. **b**, Ligands competing for the binding of ^3H -L-Glu to GluR0 S1S2 using a $1:10^5$ molar ratio of ^3H -L-Glu to cold ligand. The measurement using L-Glu was taken as 100% inhibition, and other values were normalized accordingly. Compounds were examined at 1 mM to detect low-affinity ligands. At 1 mM, L-glutamate, L-glutamine and HCA most effectively compete for ^3H -L-glutamate binding to GluR0 S1S2; L-serine, L-alanine, L-threonine, glycine, L-AA, APA and DNQX (DX) are significantly less potent. In contrast to eukaryotic GluRs, 1 mM AMPA, kainate (KAI) and NMDA (NM) do not block binding of ^3H -L-glutamate to GluR0 S1S2. Each measurement was carried out in triplicate. **c**, The K_d of ^3H -L-glutamate binding to GluR0 S1S2 is ~ 215 nM. Measurements were performed in triplicate and by filter-binding and resin-binding methods. **d**, Pharmacological profile of GluR0. IC_{50} values for selected ligands are: L-glutamate, 0.116 μM ; HCA, 1.62 μM ; L-glutamine, 16.3 μM ; L-serine, 30.0 μM ; glycine, 303 μM . IC_{50} data were from duplicate measurements. All points are given as mean $\pm$ s.e.m.

K^+ concentration in the external solution, as expected for a K^+ -selective ion channel, whereas the reversal potential of GluR2 measured under the same conditions changed much less. Analysis using the extended Goldman–Hodgkin–Katz voltage and Nernst equations gave $P_{Na}/P_K < 0.01$ for GluR0 and $P_{Na}/P_K = 0.67$ for GluR2. Because the current carried by GluR0 channels showed voltage-dependent block of inward K^+ flux by extracellular Na^+ (Fig. 4c), the membrane topology for GluR0 is inverted compared to conventional K^+ channels (Fig. 1). When the Na^+ concentration is high in the external solution, Na^+ ions become trapped in the entrance to the pore on the extracellular side of the selectivity filter. Because Na^+ is a poor substrate for the selectivity filter, K^+ flux through the channel is reduced. This idea is supported by the observation that cytoplasmic Na^+ ions in conventional K^+ channels produce voltage-dependent block of outward K^+ flux²³.

The observation of the sequence relationships between GluR0, eukaryotic GluRs and potassium channels, combined with the demonstration that GluR0 forms glutamate-gated, potassium-selective channels, enables us to make several conclusions about the structures and mechanisms of glutamate receptors and potassium channels. First, the M1, P and M2 pore-forming regions in GluRs and potassium channels have very similar architecture, with differences primarily confined to the selectivity filter region. Second, the amino-acid sequence and functional relationships between GluR0, eukaryotic GluRs and KcsA support the idea that all GluRs are tetrameric in subunit stoichiometry². Third, the prominent conservation of residues at the end of M2, which is the region of KcsA implicated in gating^{17,18}, indicates that the chemical nature of the gate and the mechanism by which it is opened and closed may be similar in all GluRs and potassium channels. Fourth,

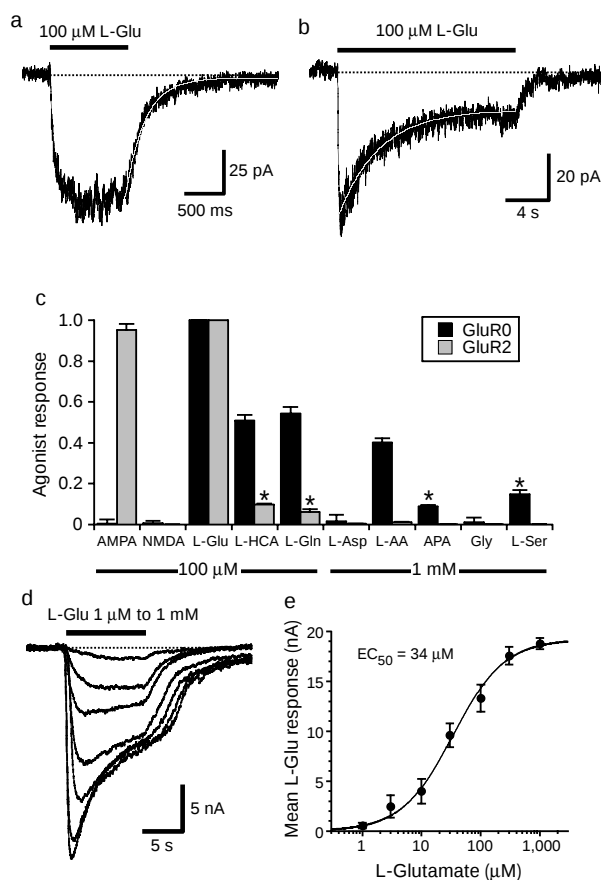


Figure 3 GluR0-mediated ion-channel responses. **a**, Slow activation and deactivation in response to rapid application of 100 μM L-glutamate to a HEK cell transiently transfected with GluR0; the 20–80% rise time was 214 ms; the line through the data points after L-glutamate removal is a least-squares-fit single exponential function of time constant 247 ms. **b**, Desensitization in response to a 17-s application of L-glutamate in the same HEK cell; the line is a single exponential function of time constant 3.82 s. **c**, Comparison of agonist sensitivity for 10 α-amino acids tested on both GluR0 and rat AMPA receptors assembled from GluR2. Responses are normalized to those evoked by 100 μM L-glutamate in *Xenopus* oocytes (5–17 cells per agonist for GluR0 and 4–8 cells for GluR2); asterisk, response > 0 , $P < 0.01$ (one sample *t*-test). **d**, GluR0 responses to 10-s applications of 1 μM to 1 mM L-glutamate recorded from a single oocyte; to allow recovery from desensitization, agonists were applied at 3-min intervals; responses to repeated applications of 100 μM L-glutamate interspersed with those to other concentrations were used to correct for rundown. **e**, Dose–response analysis for responses to L-glutamate from 7 oocytes; data points, mean $\pm$ s.d.; responses in individual oocytes were fit with a single binding-site isotherm allowing I_{max} (11.6–29.5 nA) to float. Responses are normalized to the mean value for I_{max} of 19.2 ± 6.5 nA.

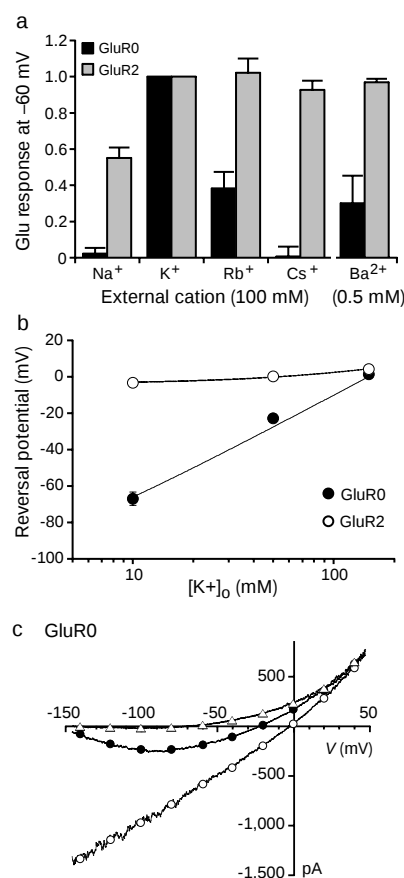


Figure 4 GluR0 gates a K^+ -selective Ba^{2+} -sensitive conductance. **a**, Responses to 100 μM L-glutamate recorded at -60 mV from *Xenopus* oocytes expressing either GluR0 or GluR2 with 100 mM Na^+ , K^+ , Rb^+ or Cs^+ as the sole extracellular monovalent cation, and with a K^+ -channel blocker, 0.5 mM Ba^{2+} , added to the solution. Error bars, s.d. for 6–7 cells per solution normalized to responses recorded with 100 mM K^+ . **b**, Reversal potential measurements reveal high K^+ selectivity for GluR0 but not GluR2. Responses are mean $\pm$ s.d. for 5–11 HEK cells tested with 10, 50 and 150 mM $[K^+]_o$ with Na^+ used to maintain constant ionic strength. **c**, GluR0 responses show voltage-dependent block by extracellular Na^+ . The current–voltage plot shows leak-subtracted responses from a single HEK cell with large GluR0 currents during ramp changes in membrane potential from -150 to +50 mV (0.8 mV ms^{-1}), recording using either 10 mM $[K^+]_o$ plus 140 mM $[Na^+]_o$ (triangles), 50 mM $[K^+]_o$ plus 100 mM $[Na^+]_o$ (filled circles) or 150 mM $[K^+]_o$ with no added Na^+ (open circles). Similar responses with a negative slope conductance at hyperpolarized membrane potentials were recorded in 11 and 9 cells with 100 or 140 mM $[Na^+]_o$, respectively; with 150 mM $[K^+]_o$, weakly outward rectifying responses were recorded from 5 cells.

the absence of M3 in GluR0 indicates that this transmembrane segment may not be essential for the structure or function of GluR ligand-gated channels, and that its presence in eukaryotic GluRs may be required to fine-tune the kinetic properties of the channel, as well as to provide an intracellular 'handle' for interaction with cytoskeletal and signal transduction molecules. Fifth, the observation that GluR0 does not have an N-terminal domain in GluR0 implies that it has a modulatory role in eukaryotic GluRs. In contrast to previous studies on kainate-binding proteins²⁴, our results show that the N-terminal domain is not essential for channel gating. Finally, the fact that GluR0 is related to eukaryotic GluRs in amino-acid sequence and biological function indicates that GluR0, or an as yet undiscovered GluR0 homologue, was the evolutionary precursor to eukaryotic, ionotropic glutamate receptors. □

Methods

Gene cloning, expression and protein purification

The full-length and S1S2 coding regions of GluR0 were cloned from *Synechocystis* PCC 6803 genomic DNA by the polymerase chain reaction (PCR) and placed into pETGQ²². The S1 fragment spanned residues 21–139; S2 included 245–397; and S1 and S2 were connected by a five-residue linker (GTDGN); the N terminus of S1 included residues MH₈SSGLVPRGSAM, which comprised the His₈ tag, the thrombin cleavage site and a short linker. The site of cleavage of the signal peptide (residues 1–19), was predicted using SignalP (<http://www.cbs.dtu.dk/services/SignalP/>). A construct in which the first 39 amino acids of GluR0 were replaced by the 31-amino-acid signal peptide of GluR6 together with 91 base pairs of GluR6 5' untranslated regions was used for physiological studies. For expression in *Xenopus* oocytes, complementary RNA was transcribed from pGEMHE²⁵. The eukaryotic expression vector for expression in HEK cells¹⁰ contained an internal ribosome entry site and the cDNA for enhanced green fluorescent protein following the coding region for GluR0.

The water-soluble GluR0 S1S2 construct was expressed in BL21(DE3) *Escherichia coli* cells following induction by 1 mM IPTG at 30 °C for 3 h. After the cells were collected by centrifugation, they were disrupted using a French Press and GluR0 S1S2 was isolated by NTA chromatography. The eluate from the Ni²⁺ column was concentrated and GluR0 S1S2 was further purified by passage over an XK 26/100 Superose 12 column in a buffer containing 20 mM Tris-Cl, pH 7.4, 200 mM NaCl and 1 mM EDTA. Protein concentration was determined spectrophotometrically ($\epsilon = 33,000 \text{ M}^{-1} \text{ cm}^{-1}$ where ϵ is the extinction coefficient at 280 nm).

Ligand binding

For ligand screening experiments, binding reactions were carried out in ligand-binding buffer (LBB; 30 mM Tris-HCl, pH 7.2, 2.5 mM CaCl₂, 100 mM KSCN, 10% glycerol²⁶) in a total volume of 500 μl at 0 °C for 1 h. The binding-reaction solution contained 250 nM protein, 1 mM cold ligand and 50 nM radiolabelled ligand. The membrane filter binding assays using PEI-treated GF/B filters and GSWP 02500 membranes were performed as described²². The recrystallized L-glutamine used in the ligand-binding and physiological experiments contained less than 0.08% L-glutamate, as determined by quantitative amino-acid analysis.

One set of K_d measurements using Ni-NTA resin was carried out in ligand-binding reaction mixtures (50 μl of Ni-NTA resin, 350 μl of 90 nM GluR0 S1S2, and the appropriate volume of a 20 μM stock of L-glutamate (³H:H 1:19)); LBB was added to bring the final volume to 500 μl . After incubation on ice (2 h), the resin was centrifuged and the supernatant removed. The tubes were centrifuged again and residual supernatant removed using a gel-loading pipette tip. The resin was then transferred to a 7-ml scintillation vial with 6 ml of scintillation liquid, and K_d measurements using filter-binding were carried out as described²².

For IC₅₀ measurements, mixtures contained 90 nM GluR0 S1S2 and different concentrations of cold ligand. The reactions were carried out at 0 °C for 1 h. The reaction mixtures were loaded onto GSWP 02500 membranes which were washed with 2 ml of ice-cold LBB, and the filters were then placed in scintillation fluid and counted. Experimental data were analysed by nonlinear curve-fitting using GraphPad Prism.

Physiology

Xenopus oocytes were injected with 50 ng GluR0 cRNA, or 20 pg cRNA for GluR2(Q)L504Y (ATG = 1), a non-desensitizing AMPA receptor²⁷, 3–7 days before recording under two-electrode voltage clamp at –60 mV using agarose-cushioned microelectrodes filled with 3 M KCl²⁸. The recording chamber had a volume of 5 μl , and solutions were applied at 250 $\mu\text{l min}^{-1}$ using computer-operated solenoid valves. Unless stated otherwise, the extracellular solution contained 100 mM KCl, 0.3 mM CaCl₂, 1 mM MgCl₂ and 5 mM HEPES, pH 7.3; in some experiments with GluR0 it contained 2 mM CaCl₂, which increased leak current stability; control experiments revealed no difference in the amplitude or kinetics of GluR0 responses at –60 mV when CaCl₂ and MgCl₂ were varied from 0.1 to 2 mM. HEK 293 cells grown on 35-mm dishes were transfected with 1.5 μg cDNA per dish for GluR0 or a 4:1 ratio of GluR2(Q)L504Y and GFP65T using calcium phosphate²⁹. Recordings were performed using whole-cell patch clamp. The external solution contained 145 mM NaCl, 5 mM KCl, 5 mM HEPES, 1 mM MgCl₂,

1.8 mM CaCl₂. Where indicated, NaCl was replaced by KCl at an equimolar ratio. The internal solution contained 105 mM K-gluconate, 15 mM KF, 15 mM KCl, 1 mM MgCl₂, 0.5 mM CaCl₂, 5 mM K₂BAPTA, 10 mM HEPES, pH 7.2 with KOH. Recordings were made using fire-polished thin-walled borosilicate glass pipettes (2–5 M Ω). Series resistance (3–10 M Ω) was routinely compensated by 90%. Agonists were applied by a stepper motor-based fast-perfusion system³⁰. Data were acquired and analysed on Power Macintosh G3 computers using 16-bit analogue-to-digital converters under the control of the data-acquisition program Synapse. Igor, Kaleidagraph and Statview were used for additional analyses.

Received 14 September; accepted 5 November 1999.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 1992).
- Dingledine, R., Borges, K., Bowie, D. & Traynelis, S. F. The glutamate receptor ion channels. *Pharmacol. Rev.* **51**, 7–61 (1999).
- Coetzee, W. A. et al. Molecular diversity of K⁺ channels. *Ann. NY Acad. Sci.* **868**, 233–285 (1999).
- Yellen, G. The bacterial K⁺ channel structure and its implications for neuronal channels. *Curr. Opin. Neurobiol.* **9**, 267–273 (1999).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Kaneko, T. et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136 (1996).
- Lomeli, H. et al. The rat delta-1 and delta-2 subunits extend the excitatory amino acid receptor family. *FEBS Lett.* **315**, 318–322 (1993).
- Lam, H. M. et al. Glutamate-receptor genes in plants. *Nature* **396**, 125–126 (1998).
- Hollmann, M., O'Shea-Greenfield, A., Rogers, S. W. & Heinemann, S. Cloning by functional expression of a member of the glutamate receptor family. *Nature* **342**, 643–648 (1989).
- Keinänen, K. et al. A family of AMPA-selective glutamate receptors. *Science* **249**, 556–560 (1990).
- Stern-Bach, Y. et al. Agonist selectivity of glutamate receptors is specified by two domains structurally related to bacterial amino acid-binding proteins. *Neuron* **13**, 1345–1357 (1994).
- Kuusinen, A., Arvola, M. & Keinänen, K. Molecular dissection of the agonist binding site of an AMPA receptor. *EMBO J.* **14**, 6327–6332 (1995).
- Armstrong, N., Sun, Y., Chen, G.-Q. & Gouaux, E. Structure of a glutamate receptor ligand binding core in complex with kainate. *Nature* **395**, 913–917 (1998).
- Lampinen, M., Pentikäinen, O., Johnson, M. S. & Keinänen, K. AMPA receptors and bacterial periplasmic amino acid-binding proteins share the ionic mechanism of ligand recognition. *EMBO J.* **17**, 4704–4711 (1998).
- Baro, D. J., Cole, C. L. & Harris-Warrick, R. M. The lobster shaw gene: cloning, sequence analysis and comparison to fly shaw. *Gene* **170**, 267–270 (1996).
- Schrempf, H. et al. A prokaryotic potassium ion channel with two predicted transmembrane segments from *Streptomyces lividans*. *EMBO J.* **14**, 5170–5178 (1995).
- Doyle, D. A. et al. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Perozo, E., Cortes, D. M. & Cuello, L. G. Structural rearrangements underlying K⁺-channel activation gating. *Science* **285**, 73–78 (1999).
- Zuo, J. et al. Neurodegeneration in Lurcher mice caused by mutation in $\delta 2$ glutamate receptor gene. *Nature* **388**, 769–773 (1997).
- Wo, Z. G. & Oswald, R. E. Unraveling the modular design of glutamate-gated ion channels. *Trends Neurosci.* **18**, 161–168 (1995).
- Wood, M. W., VanDongen, H. M. & VanDongen, A. M. Structural conservation of ion conduction pathways in K channels and glutamate receptors. *Proc. Natl Acad. Sci. USA* **92**, 4882–4886 (1995).
- Chen, G. Q. & Gouaux, E. Overexpression of a glutamate receptor (GluR2) ligand binding domain in *Escherichia coli*: Application of a novel protein folding screen. *Proc. Natl Acad. Sci. USA* **94**, 13431–13436 (1997).
- Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single *Streptomyces lividans* K⁺ channels. Functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–559 (1999).
- Wada, K. et al. Sequence and expression of a frog brain complementary DNA encoding a kainate-binding protein. *Nature* **342**, 684–689 (1989).
- Liman, E. R., Tytgat, J. & Hess, P. Subunit stoichiometry of a mammalian K⁺ channel determined by construction of multimeric cDNAs. *Neuron* **9**, 861–871 (1992).
- Keinänen, K., Köhr, G., Seeburg, P. H., Laukkanen, M.-L. & Oker-Blom, C. High-level expression of functional glutamate receptor channels in insect cells. *Bio/Technology* **12**, 802–806 (1994).
- Stern-Bach, Y., Russo, S., Neuman, M. & Rosenmund, C. A point mutation in the glutamate binding site blocks desensitization of AMPA receptors. *Neuron* **21**, 907–918 (1998).
- Schreibmayer, W., Lester, H. A. & Dascal, N. Voltage clamping of *Xenopus* oocytes utilizing agarose-cushion electrodes. *Pflügers Arch.* **426**, 453–458 (1994).
- Chen, C. & Okayama, H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7**, 2745–2752 (1987).
- Vyklicky, L., Benveniste, M. & Mayer, M. L. Modulation of N-methyl-D-aspartic acid receptor desensitization by glycine in mouse cultured hippocampal neurones. *J. Physiol. (Lond.)* **428**, 313–331 (1990).

Acknowledgements

We thank W. J. Vermaas for the *Synechocystis* PCC 6803 genomic DNA; E. Kandel for encouragement; R. MacKinnon, D. Bryant and B. Ramachandran for comments; the laboratory of D. Hirsh for the use of equipment; and C. Glasser for technical assistance. This work was supported by the Alfred P. Sloan Foundation (E.G.), the NSF Young Investigator Program (E.G.), the Klingenstein Foundation (E.G.), the National Alliance for Research on Schizophrenia and Depression (E.G.) and the NIH (E.G., M.L.M.).

Correspondence and requests for materials should be addressed to E.G. (e-mail: jeg52@columbia.edu).

Transmissible and genetic prion diseases share a common pathway of neurodegeneration

Ramanujan S. Hegde^{*†}, Patrick Tremblay^{‡§}, Darlene Groth^{‡§}, Stephen J. DeArmond^{‡§||}, Stanley B. Prusiner^{‡§¶} & Vishwanath R. Lingappa^{*‡#}

Departments of ^{*}Physiology, [‡]Institute for Neurodegenerative Diseases, [§]Neurology, ^{||}Pathology, [¶]Biochemistry and Biophysics, and [#]Medicine, University of California, San Francisco, California 94143, USA

[†] Present address: Laboratory of Cellular Oncology, National Institutes of Health, Bethesda, MD 20892, USA

Prion diseases can be infectious, sporadic and genetic^{1–4}. The infectious forms of these diseases, including bovine spongiform encephalopathy and Creutzfeldt-Jakob disease, are usually characterized by the accumulation in the brain of the transmissible pathogen, an abnormally folded isoform of the prion protein (PrP) termed PrP^{Sc}. However, certain inherited PrP mutations appear to cause neurodegeneration in the absence of PrP^{Sc} (refs 5–8), working instead by favoured synthesis of CtmPrP, a transmembrane form of PrP (ref. 9). The relationship between the neurodegeneration seen in transmissible prion diseases involving PrP^{Sc} and that associated with CtmPrP has remained unclear. Here we find that the effectiveness of accumulated PrP^{Sc} in causing neurodegenerative disease depends upon the predilection of host-encoded PrP to be made in the CtmPrP form. Furthermore, the time course of PrP^{Sc} accumulation in transmissible prion disease is followed closely by increased generation of CtmPrP. Thus, the accumulation of PrP^{Sc} appears to modulate *in trans*

the events involved in generating or metabolising CtmPrP. Together, these data suggest that the events of CtmPrP-mediated neurodegeneration may represent a common step in the pathogenesis of genetic and infectious prion diseases.

We have previously shown that certain mutations in PrP (including the A117V mutation associated with human prion disease) alter its biogenesis at the endoplasmic reticulum, causing a higher percentage of PrP molecules to be synthesized in a transmembrane form that we termed CtmPrP (ref. 9). Expression of CtmPrP-favouring mutations in transgenic mice resulted in neurodegenerative changes similar to those observed in prion disease⁹. The detection of CtmPrP (but not PrP^{Sc}) in the brains of human patients and in transgenic mice carrying CtmPrP-favouring mutations suggested that elevated CtmPrP causes neurodegeneration⁹, but whether this mechanism of neurodegeneration is also involved in the pathogenesis of transmissible prion disease has been unclear. To explore this question we first generated mice with mutant PrP transgenes differing in their propensity to form CtmPrP, and subsequently assessed their susceptibility to PrP^{Sc}-induced neurodegeneration.

Figure 1a shows the *in vitro* translocation products of four mutants of Syrian hamster prion protein (SHaPrP) that alter the amount of CtmPrP synthesized at the endoplasmic reticulum. Transgenic (Tg) mice expressing each of these mutant PrPs in the FVB/Pmp^{0/0} background (null for the endogenous PrP gene) were generated and characterized (see Table 1 and Fig. 1b). The Tg[SHaPrP(KH → II)_H], Tg[SHaPrP(KH → II)_M], Tg[SHaPrP(A117V)_H] and Tg[SHaPrP(N108I)_H] mice were observed to develop signs and symptoms of neurodegenerative disease at about 60, 472, 572 and 312 days, respectively (Fig. 1c and Table 1). By contrast, neither the Tg[SHaPrP(ΔSTE)] mice nor mice expressing lower levels of the disease-associated transgenes, Tg[SHaPrP(KH → II)_L], Tg[SHaPrP(A117V)_L] and Tg[SHaPrP(N108I)_L], developed spontaneous disease (Table 1). Biochemical

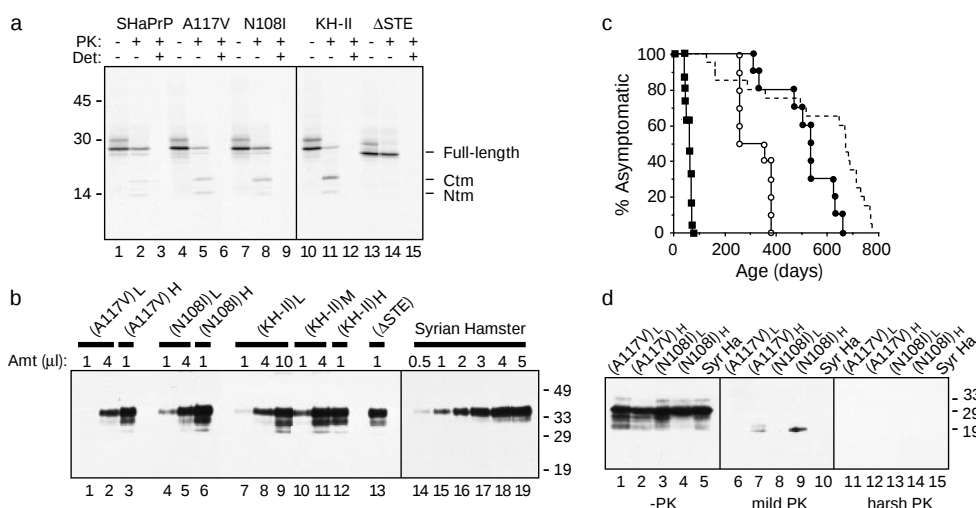


Figure 1 Dose response of CtmPrP-induced neurodegeneration. **a**, Topology of wild-type and mutant PrP molecules at the endoplasmic reticulum. *In vitro* synthesized transcript coding for each PrP construct (indicated above the gels) was used to programme a rabbit reticulocyte lysate cell-free translation reaction containing microsomal membranes and a competitive peptide inhibitor of glycosylation. Following translation, samples were either left untreated or digested with PK in the absence or presence of 0.5% Triton X-100 ('Det') as indicated above the gel. The positions of the full-length PrP molecule, the amino-terminal fragment (indicative of NtmPrP) and the carboxy-terminal fragment (indicative of CtmPrP) generated by PK digestion are marked on the right. **b**, Level of expression of various transgenic mouse lines. Varying amounts (in microlitres, indicated above the lanes) of 10% brain homogenate from each transgenic mouse were immunoblotted for

PrP with the 13A5 monoclonal antibody and compared to a titration of Syrian hamster brain homogenate. **c**, Time course of development of symptoms in Tg[SHaPrP(A117V)_H] mice (closed circles) and Tg[SHaPrP(N108I)_H] mice (open circles). Non-transgenic control animals (dashed line) and Tg[SHaPrP(KH → II)_H] mice (closed squares; data from ref. 9) are shown for comparison. **d**, Analysis of various transgenic mice and Syrian hamsters (as indicated above the gel) for CtmPrP and PrP^{Sc} as described in Methods. Tg[SHaPrP(A117V)_H] and Tg[SHaPrP(N108I)_H] samples were from clinically ill mice, and Tg[SHaPrP(A117V)_L] and Tg[SHaPrP(N108I)_L] samples were from mice (which showed no signs of disease) aged at least 600 days. The fragment of PrP resistant to PK under the 'mild' conditions, indicative of CtmPrP, is only seen in the Tg[SHaPrP(A117V)_H] and Tg[SHaPrP(N108I)_H] samples. No PK resistant PrP^{Sc} was observed in any of the samples.

analyses of brain tissue from each of these lines of transgenic mice revealed elevated $C^{tm}PrP$, but no PrP^{Sc} , in the lines which developed disease (Fig. 1d). Together, the data in Fig. 1 recapitulate the point that increased synthesis of the $C^{tm}PrP$ form of PrP is associated with the development of neurodegenerative diseases.

More remarkable is the apparent dose response relationship, seen in two ways, between $C^{tm}PrP$ and disease. First, the more strongly $C^{tm}PrP$ synthesis is favoured at the endoplasmic reticulum ($KH \rightarrow II > N108I > A117V$), the earlier is the onset of spontaneous disease (Table 1). Second, lowering the level of expression of each of these mutations below an apparent threshold stops both the generation of $C^{tm}PrP$ (Fig. 1d and ref. 9) and development of disease (Table 1). Furthermore, the three transgenic lines expressing the $KH \rightarrow II$ mutation develop disease at times inversely correlated with their respective levels of expression (Table 1). These observations demonstrate that both the $C^{tm}PrP$ -favouring quality of a mutation and its level of expression contribute to the development of neurodegeneration.

The transgenic mice with differing propensities to make $C^{tm}PrP$ were used to establish the relationship between $C^{tm}PrP$ and PrP^{Sc} . We first examined the susceptibility to PrP^{Sc} of transgenic mice with similar levels of transgene expression but differing propensities to make $C^{tm}PrP$: these mice were Tg[SHaPrP(ΔSTE)] and Tg[SHaPrP(A117V)_H]. Upon inoculation with Sc237 hamster prions, we found that the Tg[SHaPrP(ΔSTE)] and Tg[SHaPrP(A117V)_H] mice developed illness at approximately 323 and 54 days, respectively (Fig. 2a, Table 1). Biochemical analysis of representative mice at the time of disease onset revealed that the Tg[SHaPrP(ΔSTE)] mice contained substantially more PrP^{Sc} than the Tg[SHaPrP(A117V)_H] mice (Fig. 2b). Thus, the transgenic line that generates higher $C^{tm}PrP$ is more susceptible to PrP^{Sc} , developing disease at a lower level of overall PrP^{Sc} accumulation.

We next compared the susceptibility to Sc237 of Tg[SHaPrP($KH \rightarrow II$)_L] versus Tg[SHaPrP($KH \rightarrow II$)_M] mice (Fig. 2c), and Tg[SHaPrP(A117V)_L] versus Tg[SHaPrP(A117V)_H] mice (Fig. 2e). By holding the mutation constant, issues regarding a potential barrier to propagation are avoided, while still changing the propensity to generate $C^{tm}PrP$ by adjusting the level of expression. As expected, lowering the level of expression increased the incubation time to disease following Sc237 inoculation. More remarkably, however, we found that with both the $KH \rightarrow II$ and A117V mutants, the lower-level expressors contained more PrP^{Sc} at the time of disease onset (Fig. 2d and f). A similar inverse relationship between level of expression and amount of PrP^{Sc} at disease onset has been

observed in mice expressing different levels of wild-type PrP (ref. 10 and unpublished observations). Thus, as above, the mice with a diminished propensity to form $C^{tm}PrP$ had accumulated higher levels of PrP^{Sc} at the time of disease onset.

To integrate the above inoculation data into a single plot, we ranked the different lines of transgenic mice by their relative propensities to generate $C^{tm}PrP$ using a measure that we term the Ctm-index (see Table 1). This index is derived by multiplying the percentage of chains synthesized in the $C^{tm}PrP$ topology by the level of transgene expression, thus combining the two parameters known to influence $C^{tm}PrP$ generation. Figure 2g shows that a relationship exists between the Ctm-index and the amount of PrP^{Sc} that had accumulated at disease onset. The ability of PrP^{Sc} to cause disease seems to be a function of the propensity of the host to generate $C^{tm}PrP$.

The data in Fig. 2 demonstrate two important points. First, very different levels of PrP^{Sc} accumulation are observed at the time of onset of neurologic dysfunction upon inoculation of the various lines of transgenic mice. Given that the strain of the mice is identical in each case, with the only differences being in the nature and level of expression of the PrP transgene, we conclude that accumulation of protease-resistant PrP^{Sc} is not likely to be the most proximate cause of disease; subsequent events (apparently involving $C^{tm}PrP$) are likely to be involved. Second, there is a relationship between the amount of accumulated PrP^{Sc} and the factors that modulate $C^{tm}PrP$ generation. Such a relationship argues that $C^{tm}PrP$ and PrP^{Sc} are part of a pathway in which each can potentially influence, either directly or indirectly, the metabolism of the other.

One way to explain the data in Fig. 2g is if accumulation of PrP^{Sc} causes increased generation of $C^{tm}PrP$, which then elicits neurodegeneration. Thus, transgenic mice that have a greater propensity to generate $C^{tm}PrP$ (that is, a high Ctm-index) would require less PrP^{Sc} accumulation before $C^{tm}PrP$ generation is increased beyond the threshold needed to produce detectable neurologic disease. This model would explain the inverse relationship observed in Fig. 2g, and makes two additional predictions. First, as transgenic mice that substantially favour synthesis of PrP in the $C^{tm}PrP$ form can entirely circumvent the requirement for PrP^{Sc} in the development of neurodegenerative disease, tissue from such mice should not be infectious. Second, $C^{tm}PrP$ levels should increase during the course of PrP^{Sc} accumulation in infectious prion disease. These predictions were tested.

To assess the transmissibility of $C^{tm}PrP$ -associated disease, brain homogenate from clinically ill Tg[SHaPrP($KH \rightarrow II$)_H] mice was

Table 1 Characteristics of transgenic mouse lines used in this study

Transgenic line name	Transgenic line number	% Ctm <i>in vitro</i> *	Level of PrP expression (relative to SHa)†	Ctm-index‡	Age of onset of disease (days)§	$C^{tm}PrP$ <i>in vivo</i> ¶	Sc237 inoculation time (days)¶¶
TgSHaPrP(ΔSTE) _H	F1788	6	4	24			323 ± 14 (9/9)
TgSHaPrP(A117V) _L	E15781	31	0.4	12			70 ± 2 (6/6)
TgSHaPrP(A117V) _H	E15727	31	4	124	572 ± 35 (5/5)	+	55 ± 1 (6/6)
TgSHaPrP(N108I) _L	E15786	35	1	35			311 ± 3 (3/3)
TgSHaPrP(N108I) _H	E15790	35	5	175	312 ± 24 (7/7)	+	233 ± 5 (9/9)
TgSHaPrP($KH \rightarrow II$) _L	E12485	48	0.4	19			257 ± 2 (9/9)
TgSHaPrP($KH \rightarrow II$) _M	F1220	48	1	48	472 ± 13 (6/6)	+	181 ± 5 (10/10)
TgSHaPrP($KH \rightarrow II$) _H	F1198	48	4	192	58 ± 11 (24/24)#	+	ND*

* Values were derived from quantification of Fig. 1a.

† Levels of PrP expression were determined by quantitative western blotting with the 13A5 monoclonal antibody and are expressed relative to PrP expression in Syrian hamster (SHa; see Fig. 1b for a representative experiment).

‡ The Ctm-index for each transgenic line is derived by multiplying the values in the preceding two columns.

§ The age of onset of clinical symptoms [average ± s.e.m. (n/n₀)].

¶ Biochemical assay for determining the presence of $C^{tm}PrP$ *in vivo* was as previously described⁹, and carried out on either clinically ill animals (in the case of transgenic lines developing illness) or mice over 600 days of age (in the case of transgenic lines that do not develop neurodegeneration).

¶¶ Time from inoculation with Sc237 SHa prions to development of neurologic signs of dysfunction [average ± s.e.m. (n/n₀)].

Data from ref. 9.

* ND, not determined.

H, higher; M, medium; L, lower.

TgSHaPrP, transgenic Syrian hamster prion protein.

inoculated intracerebrally into four hosts: Tg[SHaPrP(KH → II)_L] mice expressing the KH → II mutation at low levels; Tg[SHaPrP] mice overexpressing wild-type SHaPrP; FVB/Prnp^{0/0} mice with a homozygous disruption of the PrP gene; and Syrian hamsters. As shown in Fig. 3, homogenate from terminally sick Tg[SHaPrP(KH → II)_H] mice did not induce neurological illness at rates that were different from those inoculated with control Tg[SHaPrP] homogenate when directly compared in three independent hosts. Furthermore, biochemical and pathological examination of representative brain tissue from animals depicted in Fig. 3 at up to 625 days after inoculation did not show any evidence of PrP^{Sc} or neurologic disease in either experimental or control animals (data not shown). However, inoculation of

Tg[SHaPrP(KH → II)_L] mice with Sc237 prions readily generated PrP^{Sc} in brain tissue (Fig. 2d), which re-transmitted disease to Tg[SHaPrP(KH → II)_L] and Tg[SHaPrP] mice (data not shown). Although PrP(KH → II) is capable of being converted into PrP^{Sc}, the CtmPrP-associated disease in Tg[SHaPrP(KH → II)_H] mice does not generate detectable PrP^{Sc} and is therefore not infectious. Lack of transmission provides further support for the hypothesis that neurodegeneration in these genetic prion diseases is caused by CtmPrP directly.

The second prediction made on the basis of the data in Fig. 2g was that accumulation of PrP^{Sc} in infectious prion disease should induce increased generation of CtmPrP, which would subsequently lead to neurodegeneration. Unfortunately, testing this prediction directly is hampered by the biochemical properties of the accumulating PrP^{Sc} (ref. 11). Being highly protease-resistant and heterogeneous in its fractionation, it tends to contaminate substantially all subcellular fractions. In addition, it interferes with the assays for CtmPrP detection, which are also based on protection from proteases. Because PrP^{Sc} is not readily degraded by the cell and accumulates to very high levels¹², even very small amounts of contamination of subcellular fractions are sufficient to make detection of slight increases in CtmPrP difficult. Thus, an alternate method is required to monitor the effect of accumulating PrP^{Sc} on the topology of newly synthesized PrP (see Fig. 4a).

To design such an experiment, we took advantage of three observations. First, a species barrier to PrP^{Sc} conversion exists between mouse and Syrian hamster^{13,14}. Second, in contrast to the species barrier for PrP^{Sc} formation, no species-specific differences in the synthesis, translocation or topology are observed between mouse prion protein (MoPrP) and SHaPrP (ref. 9 and data not shown). And, finally, monoclonal antibodies highly specific to SHaPrP (which do not cross-react with MoPrP) are available to distinguish between expression of these two PrP transgenes¹⁵. Thus, in such double-transgenic animals we can use hamster-CtmPrP formation as a 'reporter' of signalling in *trans* during the course of accumulation of mouse PrP^{Sc}. For this experiment, the double-transgenic mice which synthesize both MoPrP and SHaPrP (see Methods) are inoculated with mouse prions (of the RML strain). Then, at various intervals during the time course of accumulation of PrP^{Sc} and development of disease, individual mice are killed and

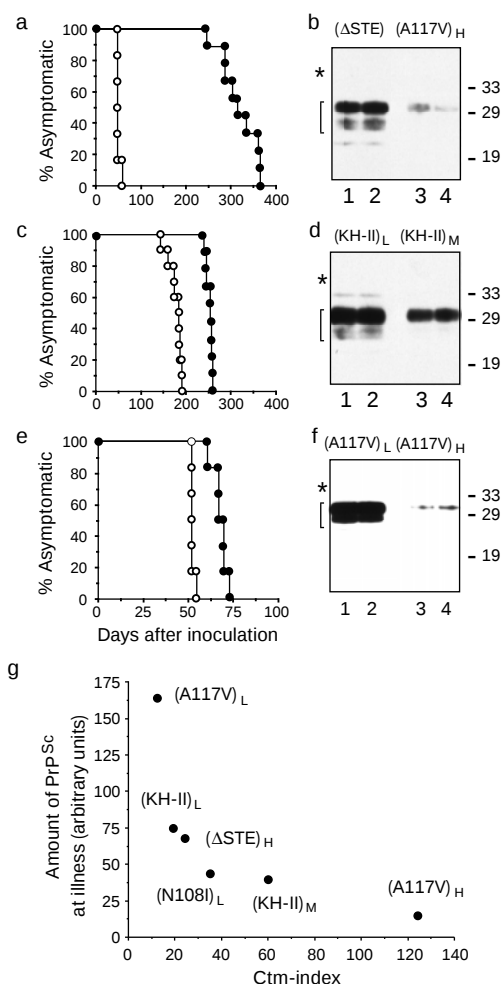


Figure 2 Relationship between CtmPrP and PrP^{Sc}. **a, c, e**, Time course of development of illness in various transgenic lines following inoculation with Sc237 hamster prions. **b, d, f**, Relative levels of protease-resistant PrP^{Sc} at time of illness in corresponding transgenic lines. Duplicate samples of each line were digested using 'harsh PK' conditions as described in Methods and equivalent amounts of each sample separated by SDS-PAGE. The C-terminal PrP27-30 fragment characteristic of PrP^{Sc} (indicated with a bracket on the left) was detected by immunoblotting with the 13A5 monoclonal antibody. The position of undigested, full-length PrP is indicated with an asterisk on the left. **g**, The Ctm-index for each transgenic line (see Table 1) was plotted against the amount of PrP^{Sc} accumulated at the time of illness following inoculation with Sc237 prions. **a**, Open circles: Tg[SHaPrP(A117V)_L]; closed circles: Tg[SHaPrP(ΔSTE)]. **c**, Open circles: Tg[SHaPrP(KH → II)_L]; closed circles: Tg[SHaPrP(KH → II)_H]. **e**, Open circles: Tg[SHaPrP(A117V)_L]; closed circles: Tg[SHaPrP(A117V)_H].

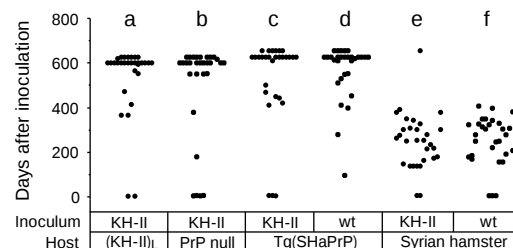


Figure 3 Lack of transmission of CtmPrP induced neurodegenerative disease. Terminally ill Tg[SHaPrP(KH → II)_H] mice ('KH → II'; **a–c, e**) and clinically normal Tg[SHaPrP] mice ('wt'; **d, f**) were sacrificed and homogenates of the brain tissue inoculated intracerebrally into various hosts as indicated below the graph. The host animals used were Tg[SHaPrP(KH → II)_L] ('(KH → II)_L'), FVB/Prnp^{0/0} ('PrP null'), Tg[SHaPrP] and Syrian hamsters. The time (in days) at which individual animals died following inoculation is plotted. Deaths by all causes, including those related to the inoculation itself, are plotted. The experiments were terminated after ~600 days with none of the remaining animals showing any signs or symptoms of neurologic disease. Each experiment represents three sets of 10 host animals injected with inoculi prepared from three separate animals to be tested. It should be noted that in our animal care facility, hamsters routinely live about 200–400 days, while mice live longer than 600 days (data not shown).

examined for total PrP^{Sc} accumulation and for the presence of hamster CtmPrP (see Fig. 4a). The principle is that, following inoculation, only MoPrP will be a substrate for prion replication and PrP^{Sc} formation¹³. The effect of this PrP^{Sc} accumulation on the ability of cells to generate (or not generate) CtmPrP can be assessed by examining SHaPrP.

Clinical disease developed in these animals about 9 weeks after inoculation (data not shown). We found that PrP^{Sc} accumulated in these mice during this 9-week time course, with the earliest times at which it was detectable being approximately 5–6 weeks (Fig. 4b). As expected, the SHaPrP was not found to have formed any PrP^{Sc} by both biochemical criteria in this study (Fig. 4b) and infectivity criteria in previous studies¹³. However, a significant increase in the amount of CtmPrP was noted upon examination of the SHaPrP (Fig. 4c). Such an increase was not observed in a parallel set of mice that did not receive the inoculum (data not shown). These findings, coupled with the observation that CtmPrP can cause neurodegeneration in the absence of a transmissible forms of PrP (ref. 9,

Figs 1 and 3), suggest that PrP^{Sc} accumulation may cause disease by inducing the synthesis of CtmPrP *de novo*.

These findings suggest causal relationships between PrP^{Sc} accumulation, the events of CtmPrP formation and metabolism, and the development of neurodegenerative disease. Three complementary and independent lines of evidence argue for this conclusion. First, increasing the generation of CtmPrP beyond a certain threshold (by modulating a combination of PrP mutation and level of expression) results in neurodegeneration in the absence of PrP^{Sc} formation (Figs 1 and 3 and ref. 9). Second, the amount of accumulated PrP^{Sc} needed to cause neurodegenerative disease is influenced by the propensity of the host to generate CtmPrP (Fig. 2). And third, the brain appears to contain increasing levels of CtmPrP during the course of accumulation of PrP^{Sc} (Fig. 4). Taken together, the data are suggestive of three successive stages in the pathogenesis of prion diseases (Fig. 5).

Infectious prion diseases are proposed to work by initiating the steps of Stage I, the accumulation of PrP^{Sc}. Genetic prion diseases could in principle work at either Stage I or II. If the PrP mutation in question results in the spontaneous formation of PrP^{Sc}, Stage I would be initiated, PrP^{Sc} would replicate and accumulate, and subsequently cause increased elevation of CtmPrP (Stage II). Such a mechanism seems plausible for the E200K mutation which causes familial Creutzfeldt-Jakob disease¹⁶. Thus, PrP^{Sc} is seen in these patients¹⁷, and the disease is readily transmissible to experimental animals⁷. Alternatively, certain other PrP mutations could bypass

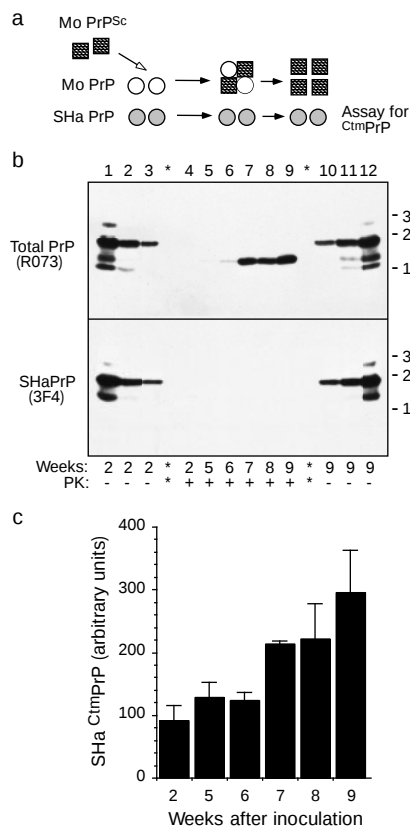


Figure 4 CtmPrP generation during time course of PrP^{Sc} accumulation. **a**, Diagram of experimental design. Double transgenic mice (see Methods) expressing both SHaPrP (shaded circles) and MoPrP (open circles) are inoculated with RML mouse prions (crosshatched squares). Over time, host MoPrP is converted to MoPrP^{Sc} and accumulates. During this time course, the SHaPrP is not converted to SHaPrP^{Sc} owing to the species barrier and may therefore be assayed for CtmPrP. **b**, Relative amounts of total PrP^{Sc} and SHaPrP^{Sc} in mice at various times (in weeks) after inoculation with RML. Homogenate was digested using the 'harsh PK' conditions (see Methods), treated with PNGase, and analysed by SDS-PAGE and immunoblotting with either the R073 polyclonal antibody (to detect total PrP) or the 3F4 monoclonal antibody (to selectively probe for SHaPrP). An equivalent amount of homogenate is analysed in each lane except lanes 2 and 11 (which contain one-fourth as much) and lanes 3 and 10 (which contain one-tenth as much). **c**, Relative amounts of Syrian hamster CtmPrP (detected selectively using the 3F4 monoclonal antibody) at various times after inoculation with RML. Each bar represents the average $\pm$ s.e.m. of three determinations.

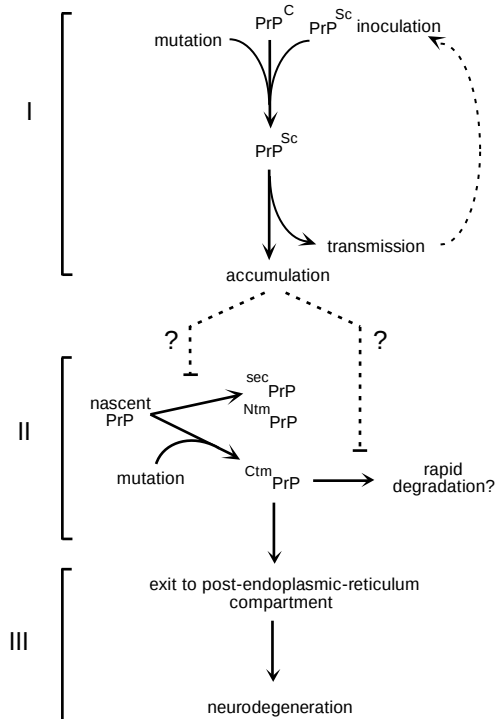


Figure 5 Three-stage model of prion disease pathogenesis. Stage I is the formation and accumulation of PrP^{Sc}. This could be initiated by either inoculation or spontaneous conversion of a mutated PrP^C to PrP^{Sc}. Stage II comprises the events involved in generating CtmPrP. These events could be affected *in trans* at a currently unknown step (dashed lines with question marks) by accumulated PrP^{Sc} or *in cis* by certain mutations within PrP. Stage III represents the events (currently unknown) involved in CtmPrP-mediated neurodegeneration. This is likely to involve exit of CtmPrP to a post endoplasmic-reticulum compartment as a first step. In this model, PrP^{Sc} can cause disease (via CtmPrP), but is not absolutely necessary. By contrast, we suggest, based on the data, that CtmPrP production is necessary and sufficient for the development of disease.

Stage I altogether by directly causing an increase in C^{tm} PrP generation. The A117V mutation resulting in human Gerstmann-Sträussler-Scheinker¹⁸ disease seems likely to work by such a mechanism. This would explain why this disease has not been transmissible^{6–8}, and why PrP^{Sc} has not been detected in the brain tissue of these patients^{6,9}.

The final stage in prion disease pathogenesis includes the mechanisms by which C^{tm} PrP, once generated, leads to neurodegenerative disease. The mechanism by which this occurs and the intracellular pathways that are involved remain entirely unclear. However, it does not appear to be the case that C^{tm} PrP is simply misfolded, retained or accumulated in the endoplasmic reticulum, or that it elicits an unfolded protein response. This is suggested by the observation that essentially all of the C^{tm} PrP has been trafficked beyond the endoplasmic reticulum (ref. 9), which is the site of the presently known quality-control machinery for protein folding in the secretory pathway^{19,20}. In addition, disease can be elicited by transgenes expressed at close to physiologic levels, as is the case with Tg[SHaPrP(KH → II)_M] animals or human cases of Gerstmann-Sträussler-Scheinker disease containing the A117V mutation. Thus, a more selective means by which C^{tm} PrP induces neurodegeneration is suggested by the available data.

The framework described in Fig. 5 proposes several new areas for future studies. First, the regulated events in C^{tm} PrP biogenesis and trafficking remain to be elucidated. The reconstitution of the early events of PrP translocation and topology determination in a cell-free system amenable to fractionation appears promising for the identification of *trans*-acting factors regulating C^{tm} PrP synthesis^{21,22}. Also, the availability of several mutants influencing topology should facilitate studies aimed at defining later events in the trafficking of C^{tm} PrP. Second, studying the metabolism of C^{tm} PrP will allow a better understanding of how these events are modulated *in trans* by PrP^{Sc} accumulation. Such studies are likely to clarify the relationship between the events of PrP^{Sc} accumulation and C^{tm} PrP-mediated neurodegeneration. Given that PrP^{Sc} accumulation may affect several metabolic functions of the cell^{23–25}, it is plausible that one or more of these influence C^{tm} PrP generation to elicit disease. The availability of PrP mutants that act at a step beyond PrP^{Sc} accumulation to directly cause C^{tm} PrP mediated neurodegeneration (ref. 9 and Fig. 1) should help to establish the downstream events in prion disease pathogenesis. □

Methods

Cell-free translation and translocation

Transcription of the relevant coding regions using SP6 polymerase, translation in rabbit reticulocyte lysate containing microsomal membranes from dog pancreas, and proteolysis were as described⁹. Translation reactions were carried out at 32 °C for 40 min, and proteolysis reactions at 0 °C for 60 min using 0.5 mg ml⁻¹ proteinase K. Products were immunoprecipitated with the R073 antibody²⁶, separated by SDS-PAGE on 15% acrylamide gels and visualized by autoradiography.

Transgenic mouse production and characterization

Transgenic mice expressing mutants (see Table 1) were generated as described (ref. 10 and references therein). PrP expression was assessed by immunoblotting of brain tissue homogenate with 13A5 mAb (ref. 27) and comparing to serial dilutions of normal Syrian hamster brain tissue (Fig. 1b and Table 1). Observation of these mice for development of spontaneous illness was as described¹³. The double transgenic mice expressing both SHaPrP and MoPrP (see Fig. 4) were generated by crossing Tg[SHaPrP]/Prnp^{0/0} (line A3922, ref. 9) to Tg[MoPrP]/Prnp^{0/0} (line B4053, ref. 28). Transmissibility (see Fig. 3) was assessed by intracerebral inoculation of 1% brain homogenate (w/v) into mice (30 µl per animal) or hamsters (50 µl per animal) as described¹³. The Sc237 strain of hamster prions (used in Fig. 2) and RML (Rocky Mountain Laboratories) strain of mouse prions (used in Fig. 4) have been described previously^{29,30}.

Assessment of brain for C^{tm} PrP and PrP^{Sc}

Brain tissue (either freshly removed or stored frozen at -80 °C following flash-freezing in liquid nitrogen) was homogenized in PBS (at 5% w/v or 10% w/v) by successive passage through 16, 18 and 20 gauge needles. For C^{tm} PrP detection ('mild' proteolysis conditions), 17-µl aliquots of the sample (at a concentration of 25 µg µl⁻¹) were adjusted (in a final

volume of 20 µl) to 1.0% NP-40, 0.25 mg ml⁻¹ PK and incubated for 60 min on ice. For PrP^{Sc} detection ('harsh' proteolysis conditions), 17 µl samples (at a concentration of 25 µg ml⁻¹) were adjusted (in a final volume of 20 µl) to 0.5% NP-40, 0.5% deoxycholate, 0.1 mg ml⁻¹ PK and incubated for 60 min at 37 °C. Note that the difference between mild versus harsh digestion conditions, while operational, is not subtle, as it involves a 37 °C change in temperature of incubation, and the presence of non-ionic detergent versus mixed micelles of non-ionic and ionic detergents. The proteolysis reactions were terminated by the addition of PMSF to 5 mM, incubating for an additional 5 min, and transferring the sample to 5 volumes of boiling 1% SDS, 0.1M Tris, pH 8.9. Samples were then digested with PNGase as directed by the manufacturer, resolved by 10% tricine-SDS-PAGE, transferred to nitrocellulose, and probed with either the 3F4 or 13A5 monoclonal antibody^{15,27} or the R073 polyclonal antibody²⁶.

Received 13 July; accepted 5 October 1999.

1. Prusiner, S. B. Prions. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
2. Weissmann, C. Molecular genetics of transmissible spongiform encephalopathies. *J. Biol. Chem.* **274**, 3–6 (1999).
3. Johnson, R. T. & Gibbs, C. J. Jr Creutzfeldt-Jakob disease and related transmissible spongiform encephalopathies. *New Eng. J. Med.* **399**, 1994–2004 (1998).
4. Horwich, A. L. & Weissman, J. S. Deadly conformations—protein misfolding in prion disease. *Cell* **89**, 499–510 (1997).
5. Brown, P. et al. Human spongiform encephalopathy: the National Institutes of Health series of 300 cases of experimentally transmitted disease. *Ann. Neurol.* **35**, 513–529 (1994).
6. Tateishi, J. et al. Immunohistochemical, molecular genetic, and transmission studies on a case of Gerstmann-Sträussler-Scheinker syndrome. *Neurology* **40**, 1578–1581 (1990).
7. Tateishi, J., Kitamoto, T., Hoque, M. Z. & Furukawa, H. Experimental transmission of Creutzfeldt-Jakob disease and related disorders to rodents. *Neurology* **46**, 532–537 (1996).
8. Tateishi, J. & Kitamoto, T. Inherited prion diseases and transmission to rodents. *Brain Path.* **5**, 53–59 (1995).
9. Hegde, R. S. et al. A transmembrane form of the prion protein in neurodegenerative disease. *Science* **279**, 827–834 (1998).
10. Manson, J. C., Clarke, A. R., McBride, P. A., McConnell, I. & Hope, J. PrP gene dosage determines the timing but not the final intensity or distribution of lesions in scrapie pathology. *Neurodegeneration* **3**, 331–340 (1994).
11. Meyer, R. K. et al. Separation and properties of cellular and scrapie prion proteins. *Proc. Natl Acad. Sci. USA* **83**, 2310–2314 (1986).
12. Czub, M., Braig, H. R. & Diringer, H. Pathogenesis of scrapie: study of the temporal development of clinical symptoms, of infectivity titres and scrapie-associated fibrils in brains of hamsters infected intraperitoneally. *J. Gen. Virol.* **67**, 2005–2009 (1986).
13. Prusiner, S. B. Transgenic studies implicate interactions between homologous PrP isoforms in scrapie prion replication. *Cell* **63**, 673–686 (1990).
14. Pattison, I. H. & Jones, K. M. Modification of a strain of mouse-adapted scrapie by passage through rats. *Res. Vet. Sci.* **9**, 408–410 (1968).
15. Kascsak, R. J. et al. Mouse polyclonal and monoclonal antibody to scrapie-associated fibril proteins. *J. Virol.* **61**, 3688–3693.
16. Spudich, S. et al. Complete penetrance of Creutzfeldt-Jakob disease in Libyan Jews carrying the E200K mutation in the prion protein gene. *Mol. Med.* **1**, 607–613 (1995).
17. Gabizon, R. et al. Insoluble wild-type and protease-resistant mutant prion protein in brains of patients with inherited prion disease. *Nature Med.* **2**, 59–64 (1996).
18. Tranchant, C. et al. Gerstmann-Sträussler-Scheinker disease in an Alsation family: clinical and genetic studies. *J. Neurol. Neurosurg. Psych.* **55**, 185–187 (1992).
19. Sidrauski, C., Chapman, R. & Walter, P. The unfolded protein response: an intracellular signalling pathway with many surprising features. *Trends Cell Biol.* **8**, 245–249 (1998).
20. Hammond, C. & Helenius, A. Quality control in the secretory pathway. *Curr. Opin. Cell Biol.* **7**, 523–529 (1995).
21. Hegde, R. S., Voigt, S. & Lingappa, V. R. Regulation of protein topology by *trans*-acting factors at the endoplasmic reticulum. *Mol. Cell* **2**, 85–91 (1998).
22. Lopez, C. D., Yost, C. S., Prusiner, S. B., Myers, R. M. & Lingappa, V. R. Unusual topogenic sequence directs prion protein biogenesis. *Science* **248**, 226–229 (1990).
23. Tatzelt, J. et al. Scrapie prions selectively modify the stress response in neuroblastoma cells. *Proc. Natl. Acad. Sci. USA* **92**, 2944–2948 (1995).
24. Wong, K. et al. Decreased receptor-mediated calcium response in prion-infected cells correlates with decreased membrane fluidity and IP₃ release. *Neurology* **47**, 741–750 (1996).
25. Johnston, A. R., Fraser, J. R., Jeffery, M. & MacLeod, N. Synaptic plasticity in the CA1 area of the hippocampus of scrapie-infected mice. *Neurobiol. Dis.* **5**, 188–195 (1998).
26. Serban, D., Taraboulos, A., DeArmond, S. J. & Prusiner, S. B. Rapid detection of Creutzfeldt-Jakob disease and scrapie prion proteins. *Neurology* **40**, 110–117 (1990).
27. Barry, R. A. & Prusiner, S. B. Monoclonal antibodies to the cellular and scrapie prion protein. *J. Infect. Dis.* **154**, 518–521 (1986).
28. Telling, G. C. et al. Interactions between wild-type and mutant prion proteins modulate neurodegeneration in transgenic mice. *Genes Dev.* **10**, 1736–1750 (1996).
29. Marsh, R. F. & Kimberlin, R. H. Comparison of scrapie and transmissible mink encephalopathy in hamsters. II. Clinical signs, pathology, and pathogenesis. *J. Infect. Dis.* **131**, 104–110 (1975).
30. Chandler, R. L. Encephalopathy in mice produced by inoculation with scrapie brain material. *Lancet* **i**, 1378–1379 (1961).

Acknowledgements

We are grateful to C. Petromilli and C. Cruz for animal care, A. Calayag and M. Zimmerman for technical assistance, J. Cayetano for assistance with neuropathology, and T. Rutkowski, N. Shah and S. Mitra for discussions and comments on this manuscript.

Correspondence and requests for materials should be addressed to V.R.L. (e-mail: vrl@itsa.ucsf.edu).

T-cell co-stimulation through B7RP-1 and ICOS

Steven K. Yoshinaga*, John S. Whoriskey*, Sanjay D. Khare*, Ulla Sarmiento*, Jane Guo*, Tom Horan*, Grace Shih*, Ming Zhang*, Marco A. Coccia*, Tadahiko Kohno*, Anna Tafuri-Bladt†, David Brankow*, Pauline Campbell*, David Chang*, Laura Chiu*, Tianang Dai*, Gordon Duncan†, Gary S. Elliott*, Ariela Hui*, Susan M. McCabe*, Sheila Scully*, Arda Shahinian†, Christine L. Shaklee*, Gwyneth Van*, Tak W. Mak† & Giorgio Senaldi*

* Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320, USA
† Amgen Research Institute, 620 University Avenue, Toronto, Ontario M5G 2C1, Canada

T-cell activation requires co-stimulation through receptors such as CD28 (refs 1–3) and antigen-specific signalling through the T-cell antigen receptor. Here we describe a new murine co-stimulatory receptor–ligand pair. The receptor, which is related to CD28 and is the homologue of the human protein ICOS⁴, is expressed on activated T cells and resting memory T cells. The ligand, which has homology to B7 molecules and is called B7-related protein-1 (B7RP-1), is expressed on B cells and macrophages. ICOS and B7RP-1 do not interact with proteins in the CD28–B7 pathway, and B7RP-1 co-stimulates T cells *in vitro* independently of CD28. Transgenic mice expressing a B7RP-1–Fc fusion protein show lymphoid hyperplasia in the spleen, lymph nodes and Peyer's patches. Presensitized mice treated with B7RP-1–Fc during antigen challenge show enhanced hypersensitivity. Therefore, B7RP-1 exhibits co-stimulatory activities *in vitro* and *in vivo*. ICOS and B7RP-1 define a new and distinct receptor–ligand pair that is structurally related to CD28–B7 and is involved in the adaptive immune response.

A full-length complementary DNA clone was isolated from a mouse cDNA library prepared from activated intestinal intra-epithelial lymphocytes (iIELs) and was found to encode a new type I transmembrane protein with 19% amino-acid identity to murine CD28. Although low, this homology is significant as the functionally related CD28 and CTLA-4 (CD152) proteins share only 26% amino-acid identity^{5–7}. Most importantly, the cysteines in this new protein (residues 42, 63, 83, 109 and 137 from the initiating methionine) are conserved that are critical for the formation of the intra- and inter-molecular disulphide bonds typical of both CD28 and CTLA-4 (ref. 8). In addition, the overall length and the relative position of the transmembrane domain are similar to those of CD28 and CTLA-4. We called this protein CD28-related protein-1 (CRP-1). An independent study has reported the identification of human ICOS⁴. CRP-1 and ICOS share 69% amino-acid identity (Fig. 1a), indicating that they are mouse and human homologues. Interestingly, murine and human CD28 have 68% homology^{5,9}.

Another full-length cDNA clone was isolated from the same iIEL library and was found to encode a new protein, B7RP-1. B7RP-1 is a type I transmembrane protein with 20% and 19% amino-acid identity to murine B7.1 (CD80) and B7.2 (CD86), respectively (Fig. 1b). This homology, like that between ICOS and CD28, is also significant, as B7.1 and B7.2 share only 24% amino-acid identity¹⁰. The homology includes the cysteines that are important for immunoglobulin loop formation at conserved locations (residues 62, 138, 185 and 242 from the initiating methionine). Moreover, the overall length and the relative position of the transmembrane domain of B7RP-1 are similar to those of the B7 molecules¹¹.

The expression of the ICOS and B7RP-1 genes in the inguinal lymph nodes from normal mice and mice subjected to the epicutaneous application of oxazolone on the abdomen was analysed by *in situ* hybridization (ISH). ICOS expression is low in the lymph

nodes from normal mice, but is readily detectable in the paracortex (a T-cell area) of those from the oxazolone-sensitized mice (Fig. 2a). This *in vivo* observation is consistent with *in vitro* observations that ICOS expression is induced on T-cells upon activation⁴. In comparison, the expression of B7RP-1 is strong in the lymph nodes from both normal and sensitized mice (Fig. 2a), particularly in the cortex (a B-cell area) and in both primary and secondary follicles. In other lymphoid tissues, B7RP-1 expression is primarily localized in B-cell areas, that is, the marginal zone of the spleen, the follicles in the Peyer's patches and the medulla of the thymus (Fig. 2b).

To determine whether ICOS and B7RP-1 interact with each other, or with CD28, CTLA-4 or B7 molecules, we performed adherent cell analysis and sorting (ACAS), which detects fluorescently labelled, soluble proteins bound to proteins expressed on adherent cells in culture. Transfected COS cells expressing the membrane-bound ICOS, B7RP-1 or CD28 proteins were incubated with purified proteins containing the human IgG1 crystallization fragment (Fc)

a					
mICOS	MKPYFCRVFV	FCFLRLTLTG	EINGSADHRM	FSFHNGGVQI	40
hICOS	MKSGLYWYFL	FCLRIKVLTV	EINGSANYEM	FIFHNGGVQI	40
	MK.....F	FC..I..LTG	EINGSA..M	F.FHNGGVQI	
mICOS	SCYPETVQQ	LKMRLFRERE	VLCELTKTGK	SGNAVSINKP	80
hICOS	LCKYPDIVQQ	FKMQLKGGQ	ILCDLTKTGK	SGNTVSIKSL	80
	.CKYP..VQQ	.KM.L....	.LC.LTRTKG	SGN.VSIK..	
mICOS	MLCLYHLSNN	SVSFFLNND	SSQGSYYFCS	LSIFDPPPFQ	120
hICOS	KFCHSQLSNN	SVSFFLYNLD	HSNANYFYCN	LSIFDPPPFK	120
	..C...LSNN	SVSFFL.N.D	.S...YYFC.	LSIFDPPPF.	
mICOS	ERNLSGGYLH	IYESQLCCQL	KLWLPVGCAG	FVVVLLFGCI	160
hICOS	-VTLTGGYLH	IYESQLCCQL	KFWLPVGCAG	FVVVCILGCI	159
	...L.GGYLH	IYESQLCCQL	K.WLP.GCAG	FVVV...GCI	
mICOS	LIWFSSKKKY	GSSVHDNPSE	YMFMAAVNTN	KKSRLAGVTS	200
hICOS	LICWLTKKKY	SSSVHDNPGE	YMFMAVNTA	KKSRLTDVTL	199
	LI.W..KKKY	.SSVHDNP.E	YMF.M.AVNT.	KKSRL...VT.	
b					
mB7RP-1	MQLKCPCFVS	LGTRQFVWKK	LHVSSGFFSG	LGLFLLLLS-	39
mCD80	MA--CNC--Q	LMQDTPL---	LKFPCCRLI-	L-LFVLLIIRL	31
	M...C.C...	L...P....	L.....	L.LF.LL...	
mB7RP-1	SLCAASAETE	VGAMVGSNVV	LSCIDPHRRH	FNLSGLYVYW	79
mCD80	SQVSSDVDEQ	LSKSVKDKVL	LPC-RYNSPH	EDESEDRIYW	70
	S.....	...V...V.	L.C.....H	S.S...YW	
mB7RP-1	QIENPEVSVT	YYLPYKSPGI	NVDSSYKNRG	HLSLDSMKQG	119
mCD80	QKHKDVV--	--LSVIAQKL	KWPEYKNR-	--TL--YDNT	100
	Q.....V...	..L.....	.V...YKNR.	...L.....	
mB7RP-1	NFSLYLKNVT	PQDTQEFTCR	VFMNTATELV	KILEEVRRLR	159
mCD80	TYSLIILGLV	LSDRGTYSQV	VQKRGRTYE	VKHLALVKLS	140
	..SL.....	..D....C.	V.....V.L.		
mB7RP-1	VAANFSTFVI	STSDSSNPGQ	ERTYTCMSKN	GYPEPNLYWI	199
mCD80	IKADFSTFNI	TESGNPSADT	KRI-TCFASG	GFPKPRFSWL	179
	..A.FSTP.I	..S.....	.R...TC...	G.P.P...W.	
mB7RP-1	-NTTDSNLID	TALQNNTVYL	NKLGLYDVIS	TLRLPWTSGR	238
mCD80	ENGRELPGIN	TTISQDPESE	LYTISSQLDF	NTTRNHTIKC	219
	.N.....I.	T.....	T...		
mB7RP-1	DVLCCVENVA	LHQNITSISQ	AESFTGNNTK	NPQETHNNEL	278
mCD80	LKYGDAHVS	EDFTWEKPEE	DDPDSKNTLV	LFAGFGGAVI	259
	V.	N.....			
mB7RP-1	KVLVPVLAVL	A-AAAFFSFI	IYRRTR-PHR	SYT-GPKTVQ	315
mCD80	TVVVIVVVIK	CFCKHRSCFR	RNEASRETNN	SLTFGPPEAL	299
	.V.V.V....	F.	R....	S.T.GP....	
mB7RP-1	LELTDDHA				322
mCD80	AEQTVFL				306
	.E.T...				

Figure 1 Amino-acid sequences of ICOS and B7RP-1. **a**, Murine ICOS aligned with human ICOS. **b**, Murine B7RP-1 aligned with murine B7.1 (CD80). Putative transmembrane regions are underlined. Experimentally determined sites of signal-peptide cleavage are indicated by asterisks. Conserved cysteines are in bold.

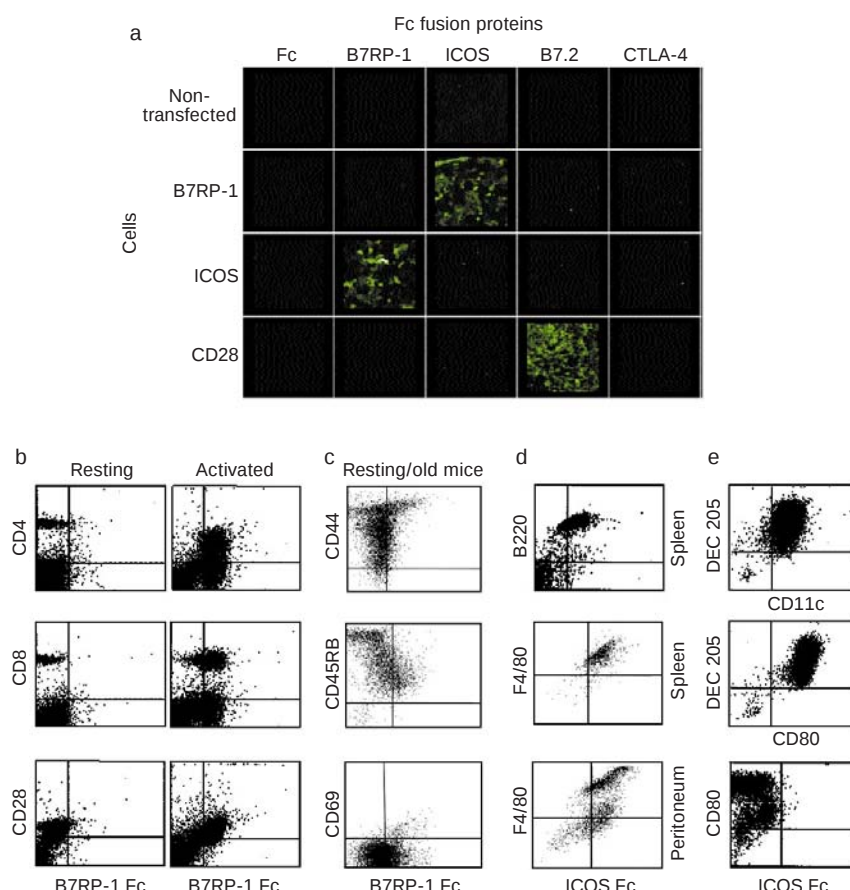
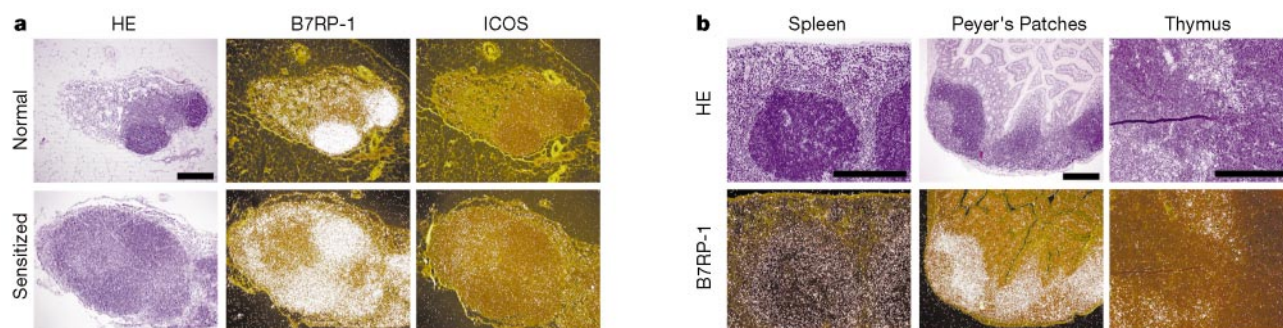


Figure 3 Interaction of ICOS with B7RP-1. **a**, Adherent cells in culture expressing membrane-bound, full-length ICOS, B7RP-1 or CD28 (left) were incubated with Fc fusion proteins (top). The cell-bound Fc fusion proteins were detected using a FITC-conjugated anti-human Fc antibody (see Methods). **b**, Suspensions of resting or PMA- and ionomycin-activated splenocytes from 10–12-week-old mice were double stained using B7RP-1–Fc labelled with an FITC-conjugated anti-human Fc antibody and a PE-conjugated antibody to CD4, CD8 or CD28. **c**, Resting splenocytes from 6–7-month-old mice were double stained using B7RP-1–Fc labelled with an FITC-conjugated anti-human Fc antibody and a PE-conjugated antibody to CD44, CD45RB or CD69. **d**, Mouse splenocytes in suspension were double stained using ICOS–Fc labelled with an FITC-conjugated anti-human Fc antibody and a PE-conjugated antibody to the B220 antigen. Macrophages, purified from suspensions of either spleen or peritoneal cells on the basis of their strong forward and

side light-scattering properties, were double stained using ICOS–Fc labelled with an FITC-conjugated anti-human Fc antibody and a PE-conjugated antibody to the F4/80 antigen (a macrophage marker). **e**, Large cells with dendritic morphology were identified on the basis of their forward and side light-scattering properties within a suspension of non-adherent cells recovered from mouse PBMC cultures. Cells were double stained using ICOS–Fc labelled with an FITC-conjugated anti-human Fc antibody and/or FITC- or PE-conjugated antibodies to CD11c, CD80 or the DEC205 antigen. PBMC cultures contained cells that were CD11c⁺/CD80⁺/DEC205⁺ but were not positively stained by ICOS-1–Fc. In **b–e**, the grid lines indicate the maximal fluorescence intensity given by control reagents, that is, non-fused Fc labelled with the FITC-conjugated anti-human Fc antibody and the FITC- or PE-conjugated isotype antibody.

fused with the extracellular domain of either ICOS, B7RP-1, B7.2 or CTLA-4. B7.2-Fc binds cells transfected with the CD28 cDNA, as expected (Fig. 3a). B7RP-1-Fc binds cells transfected with ICOS cDNA, and, conversely, ICOS-Fc binds cells transfected with B7RP-1 cDNA (Fig. 3a). These results indicate that ICOS and B7RP-1 interact in a fashion similar to CD28 and B7 molecules^{1,11}. Furthermore, B7.2-Fc does not bind cells expressing ICOS, CTLA-4-Fc does not bind cells expressing B7RP-1, and B7RP-1-Fc does not bind cells expressing CD28 (Fig. 3a). This shows that the interaction between ICOS and B7RP-1 is specific and separate from that occurring between the CD28 and B7 molecules, and indicates that ICOS and B7RP-1 represent a distinct receptor-ligand pair.

The ICOS-Fc and B7RP-1-Fc fusion proteins were used to detect cells from normal mice expressing B7RP-1 and ICOS, respectively, by fluorescence-activated cell sorting (FACS). Few CD4⁺ and virtually no CD8⁺ spleen cells from 10–12-week-old mice are posi-

tively stained by B7RP-1-Fc (Fig. 3b, top and middle left panels). After *in vitro* activation with phorbol myristate acetate (PMA) and ionomycin, however, many CD4⁺ and CD8⁺ T-cells show positive staining with B7RP-1-Fc (Fig. 3b, top and right panels). These results concur with the observation that ICOS gene expression is upregulated in lymph-node T-cell areas after immunization, and with ICOS being expressed on activated T cells⁴. Few spleen cells that stain positively for CD28 are stained by B7RP-1-Fc (Fig. 3b, bottom left panel), another indication that B7RP-1 does not interact with CD28. After activation, however, spleen cells that stain positively for CD28 are also stained by B7RP-1-Fc (Fig. 3b, bottom right panel), which reveals that ICOS and CD28 are co-expressed on the same cells after activation. To study ICOS expression on memory T cells, we collected spleen T cells from 6–7-month-old mice. Older mice have more ICOS⁺ splenic T cells than younger mice (data not shown). A conspicuous number of these ICOS⁺ cells are high in

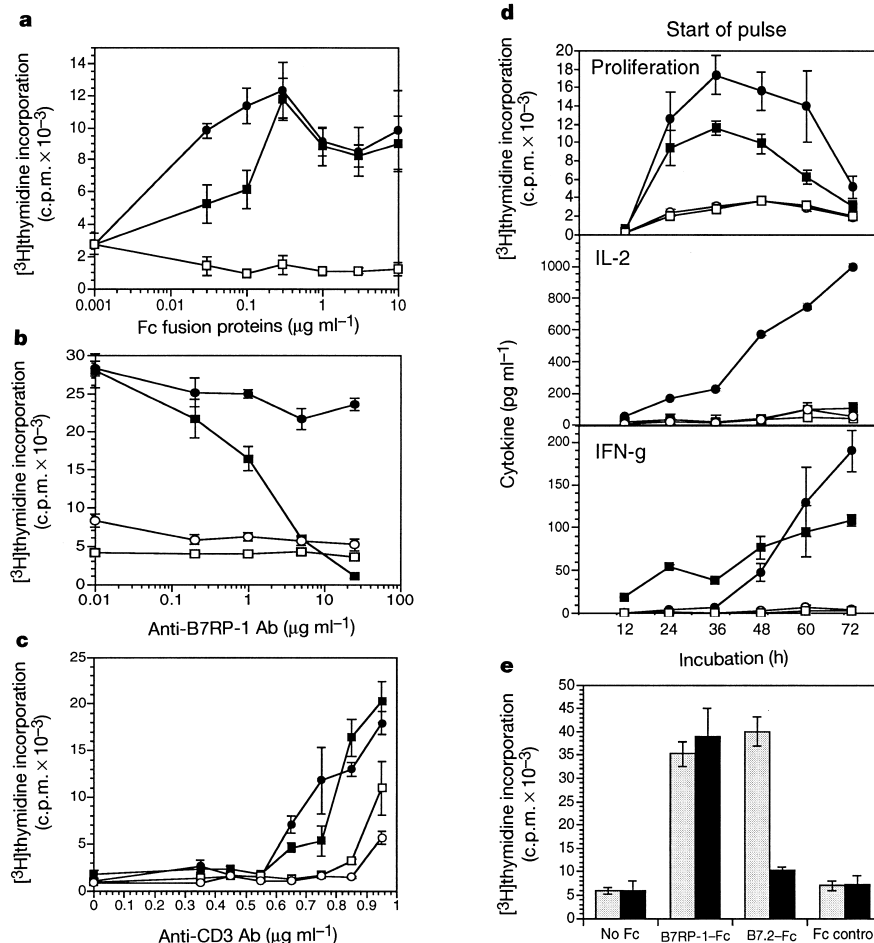


Figure 4 T-cell proliferation and cytokine secretion by B7RP-1 using B7RP-1-Fc (filled squares) and B7.2-Fc (filled circles) proteins. **a**, T-cell proliferation induced by different quantities of B7RP-1-Fc, B7.2-Fc or OPG-Fc (ref. 20) control (open squares). Fusion proteins were added at various concentrations to 96-well plates pre-coated with anti-human Fc FAb₂ (12.5 $\mu\text{g ml}^{-1}$, Sigma) and anti-CD3 antibody (0.9 $\mu\text{g ml}^{-1}$, clone 145-2C11, Pharmingen). B7RP-1-Fc and B7.2-Fc co-stimulate T-cells dose dependently to 0.3 $\mu\text{g ml}^{-1}$, when the maximal effect is achieved. **b**, T-cell proliferation induced by B7RP-1-Fc, B7.2-Fc, non-fused Fc control (open squares) or no Fc (open circles), with an anti-CD3 antibody (0.85 $\mu\text{g ml}^{-1}$) and with various concentrations of a rabbit anti-B7RP-1-Fc polyclonal antibody. The Fc fusion properties (0.3 $\mu\text{g ml}^{-1}$) were bound to the plates as above. Anti-B7RP-1 was raised against purified B7RP-1 by subcutaneous injections of antigen emulsified in adjuvant and affinity purified. The antibodies were incubated for 30 min with the Fc fusion proteins before the addition of the cells. The anti-B7RP-1 antibody specifically inhibits the T-cell proliferation induced by B7RP-1-Fc dose

dependently. **c**, T-cell proliferation, indicated by thymidine incorporation, induced by B7RP-1-Fc and B7.2-Fc at 10 $\mu\text{g ml}^{-1}$, and non-fused Fc control (open squares) or no Fc (open circles), with different quantities of anti-CD3 antibody. As with B7.2-Fc, B7RP-1-Fc induced maximal proliferation relative to the control (non-fused Fc) when the anti-CD3 antibody was used at subliminal quantities. **d**, Time course of T-cell proliferation (top), IL-2 (middle) and IFN- γ (bottom) secretion induced by B7RP-1-Fc, B7.2-Fc, non-fused Fc control (open squares) or no Fc (open circles), with the anti-CD3 antibody (0.85 $\mu\text{g ml}^{-1}$). The start of the pulse at the top of the panel designates the beginning of each 12 h thymidine incorporation period. B7RP-1-Fc induces the secretion of IFN- γ but not that of IL-2, whereas B7.2-Fc induces the secretion of large quantities of IL-2. **e**, Proliferation of T cells from CD28-deficient (filled bars) or normal (open bars) mice induced by B7RP-1-Fc and B7.2-Fc at 1 $\mu\text{g ml}^{-1}$ and non-fused Fc control or no Fc, with the anti-CD3 antibody (0.85 $\mu\text{g ml}^{-1}$). All of the above analyses were done in quadruplicate.

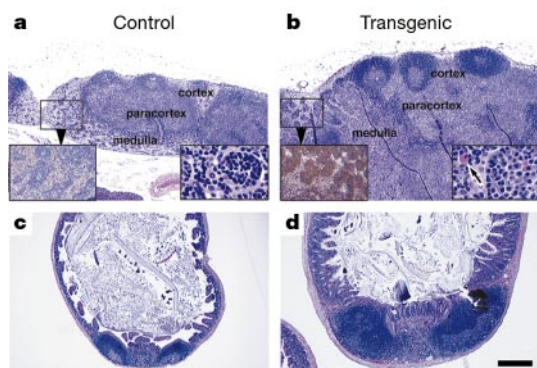


Figure 5 Mesenteric lymph nodes and Peyer's patches from B7RP-1-Fc transgenic mice and control littermates. **a, c**, Control mice (HE stain); **b, d**, transgenic mice (HE stain). In **a** and **b**, right insets are HE stains, and left insets are immunohistochemical stains for IgG. The transgenic mice have enlarged lymph nodes (**b**) with reactive hyperplasia involving the cortex, paracortex and medulla. The prominent medullary cords (**b**, right inset) comprise IgG-positive plasma cells (**b**, left inset) and scattered Russell bodies (arrows). The enlarged Peyer's patch of the transgenic mice (**d**) also shows lymphoid hyperplasia of both follicular and interfollicular tissue. Scale bar, 250 μ m.

CD44 and low in CD45RB, a profile typical of memory T cells^{12,13} (Fig. 3c, top and middle panels). These ICOS⁺ memory T cells appear to be in a resting state, as few express the activation marker CD69 (Fig. 3c, bottom panel). Therefore, activated CD4⁺ and CD8⁺ T cells and resting memory T cells appear to express ICOS.

Spleen cells expressing the B220 antigen (CD45R), a B-cell marker, are stained by ICOS-Fc (Fig. 3d, top panel), consistent with the finding that B7RP-1 gene expression is localized in the B-cell areas of lymphoid tissues. ICOS-Fc also stains splenic (Fig. 3d, middle panel) and peritoneal (Fig. 3d, bottom panel) macrophages. Dendritic cells (DCs) were generated *in vitro* from peripheral blood mononuclear cells (PBMCs) that express the DC markers DEC205 and CD11c¹⁴ (Fig. 3e, top panel). These cells also express B7.1 (Fig. 3e, middle panel) but are not stained by ICOS-Fc (Fig. 3e, bottom panel). The expression of B7RP-1 on other immune cells and dendritic cell types awaits further analysis.

To determine whether B7RP-1 has functional relevance to T cells, we incubated CD3⁺ T cells with the B7RP-1-Fc fusion protein and an anti-CD3 antibody in an *in vitro* proliferation assay. B7RP-1-Fc co-stimulates T cells to proliferate in a dose-dependent fashion (Fig. 4a), and an anti-B7RP-1-Fc antibody specifically inhibits this co-stimulation dose dependently (Fig. 4b). B7RP-1-Fc requires CD3 activation to stimulate T cells; it induces maximal T-cell proliferation relative to the control Fc protein at subliminal quantities of the anti-CD3 antibody (Fig. 4c). Under these experimental conditions, the effects of B7.2-Fc on T-cell proliferation are similar to those of B7RP-1-Fc (Fig. 4a–c).

The ability of B7RP-1-Fc to induce T cells to secrete interleukin (IL)-2 and interferon (IFN)- γ was measured in the culture supernatants of T cells under *in vitro* proliferation conditions. B7RP-1-Fc induces the secretion of IFN- γ but not IL-2, whereas B7.2-Fc induces the secretion of both IL-2 and IFN- γ (Fig. 4d). These initial observations of cytokine secretion are consistent with those made using an anti-ICOS antibody⁴.

We tested the ability of B7RP-1-Fc to co-stimulate the proliferation of T cells from CD28-deficient mice *in vitro*. B7RP-1-Fc co-stimulates T cells from either CD28-deficient or normal mice, whereas B7.2-Fc has no effect on T cells lacking CD28 (Fig. 4e). This finding shows that B7RP-1-Fc can regulate T-cell function independently of CD28 and confirms, from a functional standpoint, that the interaction between ICOS and B7RP-1 is specific and separate from that occurring between CD28 and B7.

To determine whether B7RP-1 has activity *in vivo*, we generated

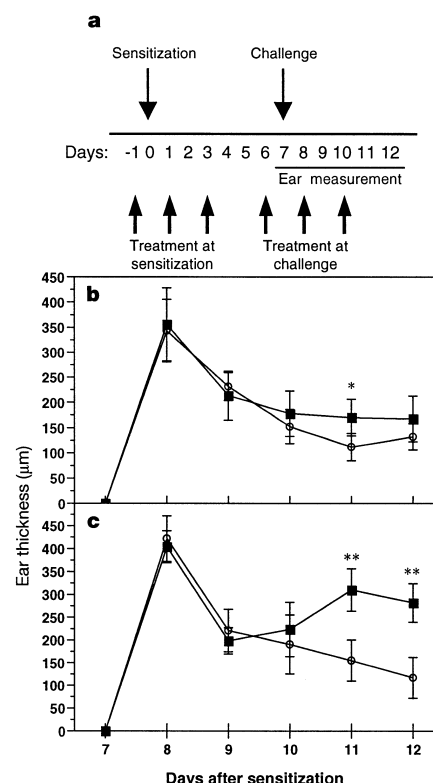


Figure 6 Contact hypersensitivity. Scheme of contact hypersensitivity induction and treatment (**a**), and effects of B7RP-1-Fc treatment during sensitization (**b**) or challenge (**c**) on contact hypersensitivity as assessed by calculating the difference in thickness between the right (challenged) and left (mock-challenged) ears. Mice treated with B7RP-1-Fc (squares) have thicker ears than saline-treated controls (circles), especially when treated at the challenge rather than the sensitization time. Asterisk, $P < 0.001$; double asterisk, $P < 0.00001$, *t*-test.

transgenic mice that express soluble B7RP-1-Fc and secrete the protein into the bloodstream¹⁵. B7RP-1-Fc transgenic mice killed at the age of 12 weeks have a distinct phenotype consisting of lymphoid hyperplasia in the spleen, lymph nodes and Peyer's patches. This hyperplasia involves both the B- and T-cell areas of the lymphoid organs (Fig. 5). This includes plasmacytosis (Fig. 5) and is accompanied by high levels of circulating IgG (mean $\pm$ s.d., 597 ± 298 mg ml⁻¹ in transgenic mice, versus 209 ± 80 mg ml⁻¹ in control littermates, $n = 7$, $P < 0.05$, *t*-test), in particular IgG2a (217 ± 100 mg ml⁻¹, versus 75 ± 29 mg ml⁻¹, $n = 7$, $P < 0.01$, *t*-test). Thus, B7RP-1 induces B- and T-cell proliferation and stimulates B cells to differentiate into plasma cells and to produce immunoglobulin. This agrees with both our observation that B7RP-1 stimulates T-cell proliferation *in vitro* and previous *in vitro* observations that an anti-ICOS antibody stimulates T-cell/B-cell cooperation⁴. These 2–3-month-old B7RP-1-Fc transgenic mice show no other alterations at necropsy, which suggests that the B7RP-1 effects are confined to immune cells.

To determine the ability of B7RP-1 to modulate the immune response *in vivo*, the B7RP-1-Fc fusion protein was administered to mice sensitized and challenged with oxazolone to induce contact hypersensitivity¹⁶. B7RP-1-Fc treatment was given at the time of either sensitization (induction of the primary immune response) or challenge (induction of the secondary immune response) (Fig. 6a). B7RP-1-Fc exacerbates contact hypersensitivity in both cases, but more notably when administered around the time of challenge than at the time of sensitization (Fig. 6b, c). This agrees with observations that B7RP-1 co-stimulates T-cell function and indicates that the T cells targeted by B7RP-1 are those involved in the secondary

immune response. This finding also shows that manipulating the ICOS–B7RP-1 interaction can modulate the secondary immune response, and thus may be a new point of immunotherapeutic intervention.

The ICOS–B7RP-1 receptor–ligand pair defines a new T-cell co-stimulatory pathway that may have several regulatory functions within adaptive immunity. ICOS–B7RP-1 may be involved in the modulation of the primary immune response; however, the role of ICOS–B7RP-1 during the primary response is probably after initial antigen presentation, because ICOS is not expressed on resting naïve T-cells. The ICOS–B7RP-1 interaction could modulate the secondary immune response by co-stimulating memory T-cell function. This is supported by the findings that ICOS is expressed on resting memory T cells, and that treatment with B7RP-1–Fc during antigen challenge results in enhanced hypersensitivity. This role in the secondary response may establish a functional difference between the ICOS–B7RP-1 pathway and other co-stimulatory pathways, such as CD28–B7 and CD40L–CD40. In fact, a blockade of either the CD28–B7 pathway or the CD40L–CD40 pathway inhibits contact hypersensitivity at the sensitization but not the challenge phase^{17,18}. Given the expression of ICOS on T cells and B7RP-1 on B cells and *in vitro* studies using an anti-ICOS antibody⁴, the ICOS–B7RP-1 interaction may also modulate T-cell-dependent antibody responses. This is strongly supported by our observation that transgenic mice expressing B7RP-1–Fc have a phenotype dominated by B-cell hyperplasia with plasmacytosis and hypergammaglobulinaemia. This study establishes ICOS and B7RP-1 as a receptor–ligand pair capable of significant co-stimulatory activity *in vitro* and *in vivo*. Future studies will be aimed at clarifying the specific roles of this pathway in the context of the adaptive immune response. □

Methods

Isolation of ICOS and B7RP-1 cDNA clones

Intestinal intra-epithelial lymphocytes were obtained from the small intestine of C57Bl/6 mice by disruption of the epithelial layer, separation on a three-step Percoll gradient, labelling with an anti-CD103 antibody (Clone M29, Pharmingen) and FACS. The RNA was prepared using the Trizol reagent (Gibco-BRL) from resting cells or cells activated overnight by plate-bound antibodies (all from Pharmingen) against γ TCR (clone GL-3), α TCR (clone H57-597) or CD3 (clone 145-2C11). Multiple preparations of RNA were pooled to construct the expressed sequence tag cDNA libraries. The DNA sequences from randomly picked cDNA clones were determined and the protein sequences translated. The deduced amino-acid sequences were then compared with those of known proteins in the public protein sequence databases.

Expression of ICOS and B7RP-1 as Fc fusion proteins

For ICOS–Fc, the sequence encoding the amino-terminal 147 amino-acids of ICOS was fused in-frame to the sequence encoding the carboxy-terminal 235 amino acids of the human IgG1 Fc and was ligated within pcDNA3 (pcDNA3/ICOS–Fc). For B7RP-1–Fc, the sequence encoding the first N-terminal 269 amino acids of B7RP-1 was fused to the human Fc gene as above (pcDNA3/B7RP-1–Fc). The coding sequences of both ICOS and B7RP-1 contained the sequences from the N terminus of each protein up to, but not including, the putative transmembrane region. 293T cells were transfected with either pcDNA3/ICOS–Fc or pcDNA3/B7RP-1–Fc, using the FuGene 6 transfection reagent (Roche Molecular Biochemicals). The Fc fusion proteins were purified from the cell supernatants using Protein A chromatography.

Adherent cell analysis and sorting

Using the FuGene 6 reagent, COS-7 cells were transiently transfected with expression vectors containing pcDNA3 and either the full-length B7RP-1 cDNA or the full-length ICOS cDNA, and plated at 5×10^5 cells per well in a 24-well plate. Chinese hamster ovary D cells, stably transfected with a pDSR α expression plasmid¹⁹ containing the cDNA for human CD28, were also plated as above. After a 48-h incubation, the purified Fc fusion proteins (ICOS–Fc, B7RP-1–Fc, B7.2–Fc and CTLA-4–Fc) were added to the cells and incubated at 37°C for 30 min. Protein–protein interactions at the cell surface were detected with an FITC-conjugated anti-human Fc antibody (Chemicon) and the ACAS Ultima Image analysis system (Meridian Instruments).

Dendritic cell preparation

Dendritic cells were generated from mouse PBMCs¹⁴. PBMCs were purified by gradient centrifugation on Histopaque-1083 (Sigma) and were plated at 5×10^6 cells ml^{−1} in tissue-culture flasks containing Iscove's complete medium supplemented with non-

essential amino acids, glutamine, sodium pyruvate, (Gibco-BRL), 2% Nutridoma SP (Boehringer–Mannheim) and 1% FCS (IN2F1). After 2-h incubation at 37°C, the non-adherent cells were removed. The remaining adherent cells were cultured in IN2F1 with rmGM-CSF, rmlL-4 and, from day 7, rmTNF (R&D Systems). Non-adherent cells were harvested for FACS on day 10.

T-cell proliferation and cytokine secretion assays

T cells from the spleens of C57Bl/6 mice were purified by negative selection through a murine T-cell enrichment column (R&D Systems). T cells (1×10^5 per well) were cultured in the absence or presence of various Fc fusion proteins for 48 h, pulsed during the last 18 h with 1 μ Ci of [³H]thymidine and harvested to count the incorporated radioactivity. T cells from the spleens of CD28-deficient mice of the C57Bl/6 strain were purified using the enrichment column as above. IL-2 and IFN- γ were measured by EIA (Biosource International) in the supernatants of T cells, which had been selected using the enrichment column described above and cultured in the presence of an anti-CD3 antibody and either B7.2–Fc, B7RP-1–Fc or no Fc fusion protein for 12, 24, 36, 48, 60 or 72 h. Supernatants from four wells were pooled and assayed in triplicate.

Preparation and analysis of transgenic mice

To target B7RP-1–Fc expression to the liver, the sequence encoding B7RP-1–Fc (see above) was subcloned into the apolipoprotein E expression vector, which is liver specific¹⁵. Micro-injection, embryo manipulation and transgene screening were conducted as described^{15,20,21}. At 12 weeks of age, nine positive transgenic mice and five negative control littermates were killed for necropsic analysis. F₁ transgenic heterozygotes and control littermates were also killed at 12 weeks of age for necropsic analysis. Organs were fixed in formalin, embedded in paraffin and stained with haematoxylin and eosin (HE) or for IgG using an immunohistochemical technique. The peripheral blood cells were counted, and the serum chemistry parameters were assessed, including the immunoglobulin classes and the IgG subclasses by EIA.

Induction and treatment of contact hypersensitivity

To induce contact hypersensitivity, 10–14-week-old female Balb/c mice were sensitized by applying a 1% solution of oxazolone (Sigma), in acetone and olive oil, onto the shaved skin of the abdomen²². Mice were challenged after 7 days by applying the oxazolone solution and the solvent control to the right and left ears, respectively. The ensuing contact hypersensitivity reaction was evaluated daily on the following 5 days by measuring the ear thickness (an indication of edema) with a micrometer, and calculating the difference between measurements of the right and left ears. B7RP-1–Fc (10 mg kg^{−1}) or saline were given intraperitoneally either the day before sensitization or the day before challenge, and twice again every 48 h (Fig. 6a).

Received 13 September; accepted 8 November 1999.

- Schweitzer, A. N. & Sharpe, A. H. The complexity of the B7–CD28/CTLA-4 costimulatory pathway. *Agents Actions* **49**, 33–43 (1998).
- Lenschow, D. J., Walunas, T. L. & Bluestone, J. A. CD28/B7 system of T cell costimulation. *Annu. Rev. Immunol.* **14**, 233–258 (1996).
- Chambers, C. A. & Allison, J. P. Co-stimulation in T cell responses. *Curr. Opin. Immunol.* **9**, 396–404 (1997).
- Hutloff, A. *et al.* ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* **397**, 263–266 (1999).
- Gross, J. A., St John, T. & Allison, J. P. The murine homologue of the T lymphocyte antigen CD28. Molecular cloning and cell surface expression. *J. Immunol.* **144**, 3201–3210 (1990).
- Brunet, J. F. *et al.* A new member of the immunoglobulin superfamily—CTLA-4. *Nature* **328**, 267–270 (1987).
- Bajorath, J. A molecular model of inducible costimulator protein and three-dimensional analysis of its reaction to the CD28 family of T cell-specific costimulatory receptors. *J. Mol. Modeling* **5**, 169–176 (1999).
- Linsley, P. S. *et al.* Binding stoichiometry of the cytotoxic T lymphocyte-associated molecule-4 (CTLA-4). A disulfide-linked homodimer binds two CD86 molecules. *J. Biol. Chem.* **270**, 15417–15424 (1995).
- Aruffo, A. & Seed, B. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl Acad. Sci. USA* **84**, 8573–8577 (1987).
- Freeman, G. J. *et al.* Murine B7-2, an alternative CTLA4 counter-receptor that costimulates T cell proliferation and interleukin 2 production. *J. Exp. Med.* **178**, 2185–2192 (1993).
- Greenfield, E. A., Nguyen, K. A. & Kuchroo, V. K. CD28/B7 costimulation: a review. *Crit. Rev. Immunol.* **18**, 389–418 (1998).
- Dutton, R. W., Bradley, L. M. & Swain, S. L. T cell memory. *Annu. Rev. Immunol.* **16**, 201–223 (1998).
- Beverly, P. C. Generation of T-cell memory. *Curr. Opin. Immunol.* **8**, 327–330 (1996).
- Inaba, K. *et al.* Identification of proliferating dendritic cell precursors in mouse blood. *J. Exp. Med.* **175**, 1157–1167 (1992).
- Simonet, W. S. *et al.* Long-term impaired neutrophil migration in mice overexpressing human interleukin-8. *J. Clin. Invest.* **94**, 1310–1319 (1994).
- Grabbe, S. & Schwarz, T. Immunoregulatory mechanisms involved in elicitation of allergic contact hypersensitivity. *Immunol. Today* **19**, 37–44 (1998).
- Tang, A., Judge, T. A. & Turka, L. A. Blockade of CD40–CD40 ligand pathway induces tolerance in murine contact hypersensitivity. *Eur. J. Immunol.* **27**, 3143–3150 (1997).
- Tang, A., Judge, T. A., Nickoloff, B. J. & Turka, L. A. Suppression of murine allergic contact dermatitis by CTLA4lg—tolerance induction of Th2 responses requires additional blockade of CD40–ligand. *J. Immunol.* **157**, 117–125 (1996).
- DeClerck, Y. A., Yean, T. D., Lu, H. S., Ting, J. & Langley, K. E. Inhibition of autolytic activation of interstitial procollagenase by recombinant metalloproteinase inhibitor MI/TIMP-2. *J. Biol. Chem.* **266**, 3893–3899 (1991).

20. Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* **89**, 309–319 (1997).
21. Brinster, R. L., Chen, H. Y., Trumbauer, M. E., Yagle, M. K. & Palmiter, R. D. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *Proc. Natl Acad. Sci. USA* **82**, 4438–4442 (1985).
22. McHale, J. F., Harari, O. A., Marshall, D. & Haskard, D. O. Vascular endothelial cell expression of ICAM-1 and VCAM-1 at the onset of eliciting contact hypersensitivity in mice: evidence for a dominant role of TNF- α . *J. Immunol.* **162**, 1648–1655 (1999).
23. Wilcox, J. N. Fundamental principles of *in situ* hybridization. *J. Histochem. Cytochem.* **41**, 1725–1733 (1993).

Acknowledgements

We thank the contributors to the Amgen Genome Program, in particular, F. J. Calzone, for direction and assistance during the course of this work. Special thanks to K. Christensen and C. DiPalma for assistance with the preparation and the analysis of the transgenic mice, and to B. Yoshinaga and E. Medlock for assistance with the preparation of this manuscript.

Correspondence and requests for materials should be addressed to S.K.Y. (e-mail: syoshina@amgen.com).

Functional recognition of the 3' splice site AG by the splicing factor U2AF³⁵

Shaoping Wu*, Charles M. Romfo†, Timothy W. Nilsen† & Michael R. Green*

* Howard Hughes Medical Institute, Program in Molecular Medicine, University of Massachusetts Medical Center, 373 Plantation Street, Worcester, Massachusetts 01605, USA

† Center for RNA Molecular Biology, Department of Molecular Biology and Microbiology, Case Western Reserve University, School of Medicine, 10,900 Euclid Avenue, Cleveland, Ohio 44106, USA

In metazoans, spliceosome assembly is initiated through recognition of the 5' splice site by U1 snRNP and the polypyrimidine tract by the U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor, U2AF (refs 1, 2). U2AF is a heterodimer comprising a large subunit, U2AF⁶⁵, and a small subunit, U2AF³⁵ (ref. 3). U2AF⁶⁵ directly contacts the polypyrimidine tract and is required for splicing *in vitro*⁴. In comparison, the role of U2AF³⁵ has been puzzling: U2AF³⁵ is highly conserved^{5–7} and is required for viability^{6,7}, but can be dispensed with for splicing *in vitro*^{4,8,9}. Here we use site-specific crosslinking to show that very early during spliceosome assembly U2AF³⁵ directly contacts the 3' splice site. Mutational analysis and *in vitro* genetic selection indicate that U2AF³⁵ has a sequence-specific RNA-binding activity that recognizes the 3'-splice-site consensus, AG/G. We show that for introns with weak polypyrimidine tracts, the U2AF³⁵–3'-splice-site interaction is critical for U2AF binding and splicing. Our results demonstrate a new biochemical activity of U2AF³⁵, identify the factor that initially recognizes the 3' splice site, and explain why the AG dinucleotide is required for the first step of splicing for some but not all introns.

The conserved AG dinucleotide at the 3' splice site is required for the second step of splicing. For some introns, called 'AG dependent', the AG dinucleotide is also essential for the earliest step(s) in spliceosome assembly and the first step of splicing (refs 10, 11, and our unpublished data). The existence of AG-dependent introns strongly suggests that a factor must recognize this sequence element during the initiation of spliceosome assembly. To identify such a putative AG-recognition factor, we prepared an adenovirus major late (Ad ML) pre-messenger RNA derivative (MINX) containing a single radioactive phosphate within the AG dinucleotide (A³²P-G) (Fig. 1a). At various times, aliquots of the splicing reaction mixture were irradiated with UV light and treated with RNase A. ³²P-tagged polypeptides were detected by SDS–polyacrylamide gel electro-

phoresis (SDS–PAGE) and autoradiography. Figure 1a shows that after a 10-min incubation in nuclear extract of HeLa cells under standard splicing conditions, 33K (33,000 relative molecular mass) and 35K polypeptides were crosslinked to the AG dinucleotide (lane 1). Intriguingly, the 35K crosslinked polypeptide was not detected in a nuclear extract depleted of U2AF (lane 2) and was restored by addition of purified U2AF heterodimer from HeLa cells (lane 3). To confirm that the 35K crosslinked polypeptide was, in fact, U2AF³⁵, we carried out UV crosslinking/immunoprecipitation experiments using anti-U2AF antibodies. Figure 1b (lanes 2 and 3) shows that the 35K crosslinked polypeptide was immunoprecipitated by either the anti-U2AF³⁵ antibody (lane 2) or an anti-U2AF⁶⁵ monoclonal antibody MC3 (lane 3).

Thus, in a HeLa nuclear extract U2AF³⁵ contacts the AG dinucleotide. We then asked whether this was an intrinsic property of U2AF or whether U2AF required other factors from the nuclear extract. Figure 1c shows that the 35K polypeptide was also crosslinked to the AG dinucleotide with both purified U2AF from HeLa cells and a baculovirus-derived U2AF⁶⁵–U2AF³⁵ heterodimer.

The U2AF³⁵–AG interaction could either result from sequence-specific RNA binding of U2AF³⁵ or be imposed in a spatially

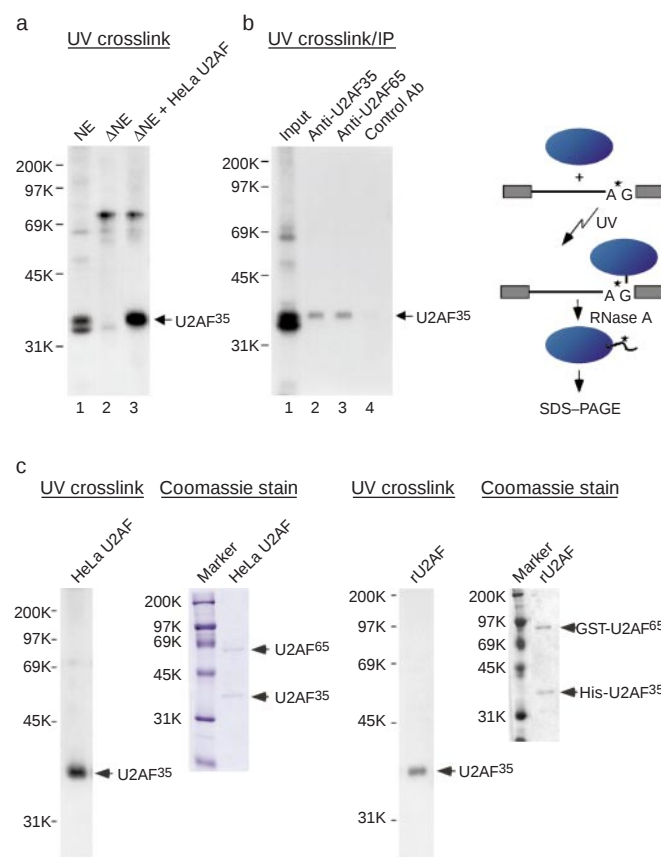


Figure 1 Detection of U2AF³⁵–AG-dinucleotide interaction by site-specific labelling and crosslinking. **a**, UV crosslinking in nuclear extracts. An A³²P site-specifically labelled adenovirus major late pre-mRNA substrate was incubated for 10 min under splicing conditions in the extracts shown and analysed by UV crosslinking as described²⁰. NE, nuclear extract; ΔNE, U2AF-depleted nuclear extract⁸; HeLa U2AF, U2AF heterodimer purified from nuclear extracts of the HeLa cells³. **b**, UV crosslinking/immunoprecipitations (IP) in nuclear extracts. As in **a** except that immunoprecipitates formed with the indicated antibodies were analysed. The U2AF⁶⁵ monoclonal antibody (MC3) was provided by M. Carmo-Fonseca²¹ and the U2AF³⁵ polyclonal antibody was provided by T. Maniatis¹⁹. **c**, UV crosslinking with purified U2AF. As in **a** except that the U2AF heterodimer was purified from HeLa cell nuclear extract or baculovirus-infected Sf9 cells (rU2AF). Polypeptides in the purified protein preparations were visualized by Coomassie blue stain.



poly(Y)-tract substitution mutant, whereas the U2AF heterodimer supported splicing of both substrates (Fig. 5b).

We have shown that U2AF³⁵ interacts with the 3' splice site in a sequence-specific fashion, and that this interaction is required for efficient U2AF binding and splicing of introns with weak poly(Y)-tracts. Introns can be divided into two classes on the basis of their requirement for the AG dinucleotide during the first step of splicing: in general, AG-dependent introns have weak poly(Y) tracts, whereas AG-independent introns have strong poly(Y) tracts¹¹. *Trans*-splicing substrates can be viewed as an extreme example of AG-dependent introns. Our results provide a molecular explanation for AG dependence. We propose that introns with strong poly(Y)-tracts efficiently bind U2AF through the large subunit and thus do not require the AG dinucleotide or U2AF⁶⁵ for the first step of splicing. In contrast, introns with weak poly(Y)-tracts have a relatively low affinity for U2AF⁶⁵, and thus require the additional contact provided by the U2AF³⁵-AG interaction for efficient U2AF binding and splicing. In this regard, *Saccharomyces cerevisiae*, in which all introns are thought to be AG independent, lacks a recognizable U2AF³⁵ homologue¹⁶, whereas *Schizosaccharomyces pombe*, in which all introns appear to be AG dependent¹⁷, has a well-conserved U2AF³⁵ orthologue⁵.

Our results, in conjunction with previous studies, indicate that U2AF contacts all three *cis*-acting elements of the 3' splice site/branch point region: U2AF⁶⁵ contacts the poly(Y)-tract through its RRM domains⁴; U2AF³⁵ contacts the 3' splice site; and the U2AF⁶⁵ (RS) domain contacts the branchpoint⁸. The first two of these contacts are sequence specific, whereas the RS domain-branchpoint interaction is spatially constrained but sequence independent. These new results underscore the pivotal role of U2AF in initial recognition of the 3'-splice-site/branchpoint region. □

Methods

Expression and purification of U2AF

The BamHI-EcoRI fragment of U2AF⁶⁵ cDNA from pGST-U2AF⁶⁵ (refs 4, 18) was inserted into the baculovirus transfer vector pAcG3X (PharMingen) to create pAcG3X-U2AF⁶⁵. Baculovirus transfer vector for U2AF³⁵ (1393-U2AF³⁵) was provided by P. Zuo and T. Maniatis¹⁹. U2AF⁶⁵ was expressed as an N-terminal GST-fusion protein and U2AF³⁵ was expressed as an N-terminal His-tagged protein. Transfer vectors containing each of the U2AF subunits were used to generate recombinant baculoviruses with a BaculoGold transfection kit (PharMingen). Virus was plaque purified and amplified to a titre of 1×10^8 PFU ml⁻¹. Sf9 insect cells were infected simultaneously with both baculoviruses at a multiplicity of infection of eight. All subsequent manipulations were carried out at 4 °C.

Cells were harvested 44 h after infection, washed and incubated in lysis buffer (50 mM Tris pH 8.0, 1 M NaCl, 0.1% Triton X-100, 5 mM β -ME and protease inhibitor cocktail (Boehringer Mannheim)). A crude extract was prepared by sonication for 2 s and centrifuged at 30,000 r.p.m. for 20 min. The soluble fraction was incubated with Ni-NTA agarose beads (Qiagen) for 20 min. The beads were loaded onto a column and washed with lysis buffer, and bound protein was eluted in lysis buffer containing 300 mM imidazole. The eluate was incubated with glutathione-agarose beads for 20 min, washed with lysis buffer and bound protein was eluted with the same buffer containing 50 mM glutathione. Peak fractions were dialysed against 100 mM KCl buffer D (20 mM HEPES pH 8.0, 0.5 mM EDTA, 20% glycerol, 1 mM DTT, 0.05% NP-40) and aliquots stored at -80 °C. Protein concentrations were estimated by comparison to BSA standards on a 10% SDS-polyacrylamide gel. Baculovirus expressed U2AF⁶⁵ was purified on glutathione-agarose beads from lysates derived from cells infected with only a U2AF⁶⁵-expressing baculovirus.

Received 19 July; accepted 5 October 1999.

1. Reed, R. Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.* **6**, 215–220 (1996).
2. Reed, R. & Palandjian, L. In *Spliceosome Assembly In Eukaryotic mRNA Processing* (ed. Krainer, A.) 103–129 (IRS Press, Oxford, 1997).
3. Zamore, P. & Green, M. R. Identification, purification and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl Acad. Sci. USA* **86**, 9243–9247 (1989).
4. Zamore, P. D., Patton, J. G. & Green, M. R. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614 (1992).
5. Wentz-Hunter, K. & Potashkin, J. The small subunit of the splicing factor U2AF is conserved in fission yeast. *Nucleic Acids Res.* **24**, 1849–1854 (1996).
6. Rudner, D. Z., Kanaar, R., Breger, K. S. & Rio, D. C. Mutations in the small subunit of the *Drosophila* U2AF splicing factor cause lethality and developmental defects. *Proc. Natl Acad. Sci. USA* **93**, 10333–10337 (1996).
7. Zorio, D. A. & Blumenthal, T. U2AF³⁵ is encoded by an essential gene clustered in an operon with RRM/cyclophilin in *Caenorhabditis elegans*. *RNA* **5**, 487–494 (1999).
8. Valcarcel, J., Gaur, R. K., Singh, R. & Green, M. R. Interaction of U2AF⁶⁵ RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* **273**, 1706–1709 (1996).

9. Kan, J. L. & Green, M. R. Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. *Genes Dev.* **15**, 462–471 (1999).
10. Lamond, A. I., Konarska, M. M. & Sharp, P. A. A mutational analysis of spliceosome assembly: evidence for the splice site collaboration during spliceosome formation. *Genes Dev.* **1**, 532–543 (1987).
11. Reed, R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113–2123 (1989).
12. Singh, R., Valcarcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173–1176 (1995).
13. Senapathy, P., Shapiro, M. B. & Harris, N. L. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183**, 252–278 (1990).
14. Mount, S. M. *et al.* Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res.* **20**, 4255–4262 (1992).
15. Nilsen, T. W. In *Trans-Splicing In Eukaryotic mRNA Processing* (ed. Krainer, A.) 310–334 (IRS Press Oxford, 1997).
16. Abovich, N. & Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **2**, 403–412 (1997).
17. Romfo, C. M. & Wise, J. A. Both the polypyrimidine tract and the 3' splice site function prior to the first step of splicing in fission yeast. *Nucleic Acids Res.* **15**, 4658–4665 (1997).
18. Fleckner, J., Zhang, M., Valcarcel, J. & Green, M. R. U2AF⁶⁵ recruits a novel human dead box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11**, 1864–1872 (1997).
19. Zuo, P. & Maniatis, T. The splicing factor U2AF³⁵ mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* **10**, 1356–1368 (1996).
20. Wu, S. & Green, M. R. Identification of a human protein that recognizes the 3' splice site during the second step of pre-mRNA splicing. *EMBO J.* **16**, 4421–4432 (1997).
21. Gama-Carvalho, M. *et al.* Targeting of U2AF⁶⁵ to sites of active splicing in the nucleus. *J. Cell Biol.* **137**, 975–987 (1997).

Acknowledgements

We thank S. DuPont, J. Varcarel, R. Singh, J. Kan and P. Maroney for suggestions and reagents; R. Reed for β -globin derivatives; P. Zuo and T. Maniatis for U2AF³⁵ antibody and baculovirus U2AF³⁵ expression construct; M. Gama-Carvalho and M. Carmo-Fonseca for U2AF⁶⁵ monoclonal antibody (MC3); L. Chiang and J. Wang for nuclear extracts. We thank T. Blumenthal for communicating results before publication. This work was supported in part by grants from the NIH to M.R.G. and T.W.N. M.R.G. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to M.R.G. (e-mail: michael.green@umassmed.edu).

Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*

Diego A. R. Zorio*† & Thomas Blumenthal*

* Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, UCHSC, Denver, Colorado 80262, USA

† Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

Introns are defined by sequences that bind components of the splicing machinery. The branchpoint consensus, polypyrimidine (poly(Y)) tract, and AG at the splice boundary comprise the mammalian 3' splice site¹. Although the AG is crucial for the recognition of introns with relatively short poly(Y) tracts, which are termed 'AG-dependent introns'², the molecule responsible for AG recognition has never been identified. A key player in 3' splice site definition is the essential heterodimeric splicing factor U2AF, which facilitates the interaction of the U2 small nuclear ribonucleoprotein particle with the branch point. The U2AF subunit with a relative molecular mass (M_r , 65K) of 65,000 (U2AF⁶⁵) binds to the poly(Y) tract^{3–7}, whereas the role of the 35K subunit (U2AF³⁵)⁸ has not been clearly defined. It is not required for splicing *in vitro*⁴ but it plays a critical role *in vivo*^{9,10}. *Caenorhabditis elegans* introns have a highly conserved U₄CAG/R at their 3' splice sites instead of branch-point and poly(Y) consensus sequences¹¹. Nevertheless, *C. elegans* has U2AF (refs 10, 12). Here we show that both U2AF subunits crosslink to the 3' splice site. Our results suggest that the U2AF⁶⁵-U2AF³⁵ complex identifies the U₄CAG/R, with U2AF³⁵ being responsible for recognition of the canonical AG.

We have proposed that U2AF is the factor that recognizes

U₄CAG/R for 3'-splice-site definition¹². To test this, we mixed ³²P-end-labelled oligonucleotides with crude embryo extracts and crosslinked them by irradiation with ultraviolet. Figure 1a shows a complex pattern of labelled bands, with one band of M_r ~70K, the size expected for U2AF⁶⁵. To investigate whether this crosslinked band was U2AF⁶⁵, we performed an immunodepletion experiment with anti-U2AF⁶⁵ antibody, followed by crosslinking by ultraviolet. As shown by the Western blot in Fig. 1b and the crosslinking experiment (Fig. 1a), both U2AF⁶⁵ and the ~70K crosslinked band were almost totally removed by the antibody, indicating that it was U2AF⁶⁵.

To obtain large quantities of U2AF made in *C. elegans* we used transgenic worms overexpressing both of the U2AF subunits, with U2AF³⁵ Flag-tagged (Fig. 1c). When crude extract from the overexpressing strain BL7014 was immunodepleted with the U2AF⁶⁵ antibody, the M_r ~70K band was less intense (data not shown). Furthermore, a comparison of the ultraviolet-cross linked 70K band with the other crosslinked bands in the same lanes indicated that it was overexpressed in the BL7014 extract (Fig. 1a). From these observations we conclude that the 70K crosslinked band is U2AF⁶⁵, and that U2AF⁶⁵ binds to the RNA oligonucleotide containing the U₄CAGG consensus sequence.

We performed several immunoprecipitation experiments to determine whether any other crosslinked bands were associated with U2AF⁶⁵. When anti-U2AF⁶⁵ antibody was used to immunoprecipitate from wild-type or BL7014 extract, two proteins crosslinked to ³²P-labelled consensus oligonucleotide were

immunoprecipitated (Fig. 2a, lanes 2 and 6): U2AF⁶⁵ itself, and another that migrated to the predicted position of U2AF³⁵. To determine whether the smaller protein was in fact U2AF³⁵ we performed immunoprecipitation with anti-Flag antibody-conjugated beads. As expected, the antibody bound neither protein in N2 extract, in which U2AF³⁵ was not tagged (Fig. 2a, lane 3). However, anti-Flag antibody precipitated ³²P-labelled crosslinked bands of 65K and 35K from the BL7014 extract (Fig. 2a, lane 7), indicating that the smaller band was indeed U2AF³⁵. These results show that both U2AF⁶⁵ and U2AF³⁵ are bound to the U₄CAGG RNA oligonucleotide directly in a ternary complex.

To determine whether the RNA binding was specific to the 3'-splice-site consensus sequence, we performed identical crosslinking experiments with oligonucleotides that contained variations of the consensus sequence. One, UA₃CAGG, disrupts the string of five pyrimidines; the others alter the AG at the 3' splice site. When three of the U residues in the pyrimidine tract were changed to A residues, no crosslinking of either U2AF subunit was seen, and no labelled bands were immunoprecipitated (Fig. 2a, lanes 11–13). In contrast, when the pyrimidine tract was left intact but the AGG was changed to GAA, GGG, AAA or GAG, U2AF⁶⁵ still crosslinked to the probe, although the levels of crosslinking were reduced (Fig. 2a, lane 9, and Fig. 2b, lanes 2–5). Importantly, U2AF³⁵ did not crosslink to the mutant oligonucleotides lacking the AGG sequence (Fig. 2a, lanes 9 and 10, and Fig. 2b, lanes 2–5). These results demonstrate that U2AF³⁵ requires the AG in the context of the intact 3' splice site to crosslink to the RNA. The faint band below U2AF⁶⁵, seen most clearly when the U2AF⁶⁵ band is attenuated by eliminating the AG, is of unknown origin; it could represent a protein that interacts with the pyrimidines and immunoprecipitates with U2AF⁶⁵.

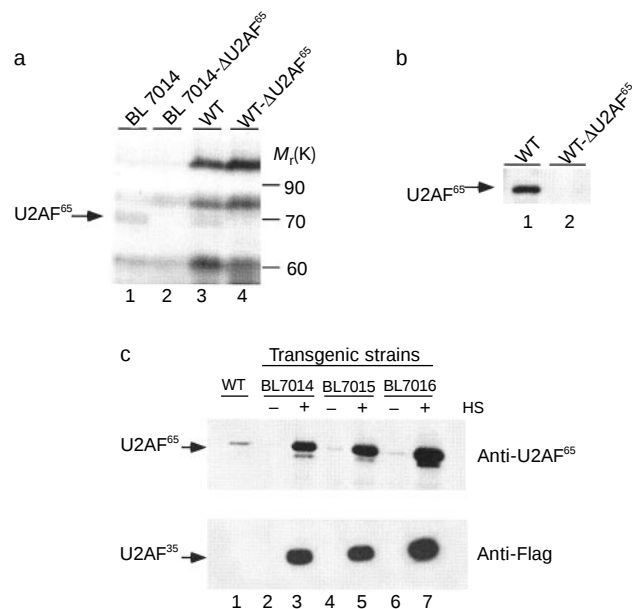


Figure 1 Crosslinking of U2AF⁶⁵ by ultraviolet irradiation of crude embryonic extracts of strains overexpressing both U2AF subunits. **a**, Identification of ultraviolet-crosslinked U2AF⁶⁵ by immunodepletion. Wild-type (WT) and BL7014 extracts were immunodepleted of U2AF⁶⁵. Equal amounts of control and U2AF⁶⁵-depleted extracts were ultraviolet-crosslinked with ³²P-labelled RNA oligonucleotide containing the 3'-splice-site consensus sequence, U₄CAGG. Samples were boiled in SDS loading buffer and subjected to electrophoresis on an SDS-8% polyacrylamide gel. The position of the band depleted by U2AF⁶⁵ antibody is indicated by an arrow (lanes 2 and 4). Note the overexpression of this band in the BL7014 extract in comparison with other crosslinked bands (lane 1). **b**, Immunoblot analysis of U2AF⁶⁵-depleted extract. Wild-type (N2) crude extract was immunodepleted of U2AF⁶⁵. Control (lane 1) and depleted (lane 2) extracts were subjected to electrophoresis, blotted and probed with U2AF⁶⁵ antibody. **c**, Western analysis of transgenic strains carrying integrated DNA arrays encoding *uaf-1* and *uaf-2* cDNAs under the control of the heat-shock promoter¹⁶. Fifty worms per lane were either heat-shocked (HS) (+) or not (-), boiled and analysed by western blotting with U2AF⁶⁵ (top) and Flag-peptide antibodies (bottom). WT, the non-transgenic wild-type control strain, N2.

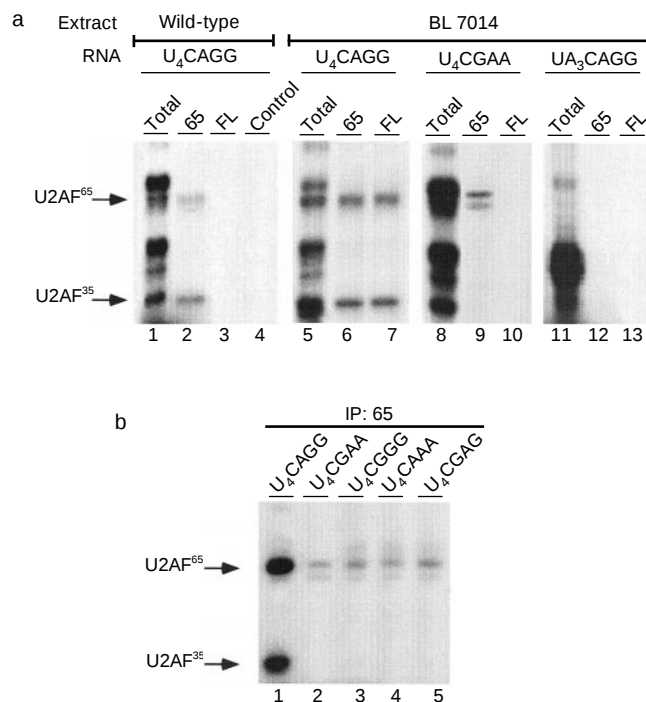


Figure 2 Both U2AF subunits crosslinked to U₄CAGG RNA oligonucleotide, but U2AF³⁵ required the AG/R consensus. **a**, Immunoprecipitation of wild-type and BL7014 extracts with various RNAs using antibodies against U2AF⁶⁵ and Flag peptide. Extracts were incubated with ³²P-labelled RNA oligonucleotides as indicated and crosslinked with ultraviolet as in Fig. 1. Wild-type or BL7014 embryo extracts were immunoprecipitated with non-immune control, U2AF⁶⁵ (65) or FLAG (FL) antibodies. 'Total', extract before immunoprecipitation. The band under U2AF⁶⁵ in lane 9 and in **b** was reproducibly observed when oligonucleotides lacking the consensus were crosslinked, but it was of unknown origin. **b**, The indicated oligonucleotides were labelled, crosslinked and immunoprecipitated (IP) with U2AF⁶⁵ antibodies (65).

Because changing the sequence of the oligonucleotide could have effects on the extent of crosslinking rather than on binding *per se*, we performed an oligonucleotide competition experiment in which only the RNA containing the consensus sequence was labelled (Fig. 3). The labelled RNA was mixed with increasing concentrations of the unlabelled oligonucleotides and crosslinked in the BL7014 embryo extract. As can be seen in Fig. 3a, U2AF⁶⁵ was specifically competed with by the U₄CAGG oligonucleotide, whereas other crosslinked proteins were not. To look specifically at the behaviour of both U2AF subunits in the competition experiment, we repeated the competition and immunoprecipitated with anti-U2AF⁶⁵ antibody (Fig. 3b, c). Again, the two oligonucleotides with the variant sequences failed to compete effectively with the labelled oligonucleotide for binding, whereas the specific oligonucleotide competed at all oligonucleotide concentrations tested. We conclude that the binding of both U2AF subunits to the 3' splice site octamer-sequence-containing oligonucleotide is specific for RNA containing the entire sequence. The two experiments combined demonstrate that the U₄CGAA oligonucleotide can bind to U2AF⁶⁵, but it does so with substantially lower affinity than does the oligonucleotide containing the consensus U₄CAGG. This suggests that U2AF³⁵ binding provides significant additional binding energy for interaction of U2AF⁶⁵ with the pyrimidine tract and is consistent with results obtained with *Drosophila* U2AF (ref. 13). In contrast,

the U₄CAGG-containing oligonucleotide does not seem to interact at all with U2AF.

Unexpectedly, the anti-Flag beads failed to immunoprecipitate the U2AF⁶⁵ labelled with the U₄CGAA oligonucleotide (Fig. 2a lane 10). It is possible that the binding of this RNA sequence alters the nature of the U2AF⁶⁵–U2AF³⁵ complex such that the U2AF⁶⁵ crosslinked to RNA is no longer immunoprecipitable by antibody against the U2AF³⁵ subunit.

C. elegans seems to have replaced the branch point and poly(Y) tract consensus sequences with the conserved octamer U₄CAG/R at the 3' splice site¹¹. Previous work has shown the importance of this sequence in intron recognition^{14,15}, but it has not previously been known what *trans*-acting factor might be responsible for interaction with this sequence. We suggested that it might be U2AF because of the similarity between the first part of the octamer and the poly(Y) tract¹². The data presented here show that both U2AF subunits are involved in RNA binding. U2AF⁶⁵ crosslinks to RNA containing the U₄C without the AG/R, whereas U2AF³⁵ crosslinks to RNA only if the complete U₄CAGR motif is present. However, the interaction of U2AF⁶⁵ with the full U₄CAGR consensus is much tighter than with the oligonucleotides from which the AG is absent. Thus, the affinity of *C. elegans* U2AF⁶⁵ for the short poly(Y) tract is significantly enhanced by an adjacent AG. It seems likely that the two-subunit protein binds to the U₄CAG/R, with U2AF⁶⁵ contacting U₄C and U2AF³⁵ contacting AG/R, although specific contacts remain to be determined. Apparently, most *C. elegans* introns have evolved an optimized binding site, U₄CAG/R, for binding the U2AF⁶⁵–U2AF³⁵ complex¹¹. Although U2AF from different organisms might recognize RNA differently, U2AF could recognize the 3' splice site in other organisms much as it does in *C. elegans*. We propose that those introns that do not require AG for the first step of splicing (AG-independent introns)² have sufficiently strong binding sites for U2AF⁶⁵ for the extra binding energy provided by the interaction of U2AF³⁵ with AG/R not to be required. Whether such complexes actually contain U2AF³⁵ *in vivo* remains to be determined. □

Methods

Construction of expression plasmids

C. elegans U2AF⁶⁵ and U2AF³⁵ were both placed under the control of the *hsp16-41* promoter¹⁶ as follows. A 1.7-kb PCR fragment containing the coding region and 3' UTR of *uaf-1* cDNA¹² was inserted into *KpnI*–*SacI*-digested pPD49.83 to create pHSUAF1, and a 911-bp fragment containing the entire *uaf-2* cDNA¹⁰ in frame with the Flag sequence at the N terminus generated by PCR was inserted into *NheI*–*SacI*-digested pPD49.83 to create pHSFLAG-UAF2.

C. elegans strains

Transgenic worm strains were generated by injecting pHSUAF1 and pHSFLAG-UAF2 together with the selection plasmid, pRF5 (ref. 17). The extrachromosomal DNA arrays were integrated by γ -irradiation (3800 rad, with a caesium source)¹⁸.

Embryonic crude extracts

N2 or BL7014 worms were grown on plates at 20 °C for 3–4 d. Worms were collected by washing with water and added to 1 litre of S-basal medium containing *E. coli* strain NA22. Cultures were incubated at 20 °C for 3–4 d with shaking at 200 r.p.m.; worms were collected by sedimentation, transferred to 50-ml polypropylene conical tubes and washed with water four or five times, followed by centrifugation in 1 M sucrose to remove debris. Worms were treated with 0.5 M NaOH and 20% bleach to isolate embryos, which were purified with 1 M sucrose and washed with sterile water. After 16 h at 20 °C, NA22 was added; the synchronous cultures were grown until the L4 stage. 5-Fluoro-2-deoxyuridine was added to 50 μ M ml⁻¹, and thymidine was added to 40 μ M ml⁻¹ to arrest growth¹⁹. The embryos contained no more than 200 cells. Embryos from strain BL7014 were heat-shocked as follows: 33 °C for 2 h, 23 °C for 30 min, 33 °C for 2 h, and recovery at 23 °C for 1 h. Heat-shocked embryos were resuspended and homogenized with a stainless steel Dounce homogenizer (Wheaton 15 ml tissue grinder) in three times the original volume of homogenization buffer (10 mM Tris–HCl, pH 8.0, 1.5 mM MgCl₂, 10 mM KCl, 1 mM dithiothreitol (DTT), 50 mM sucrose, 0.05% Nonidet P40 and one tablet (per 50 ml) of Boehringer Mannheim protease inhibitor (Complete)). The homogenate was sedimented at 8,000 r.p.m., 7,670g, and the supernatant was dialysed twice against 50 mM KCl, 1 mM DTT, 0.2 mM EDTA, 10% glycerol, 20 mM Tris–HCl, pH 8.0, and one tablet of Complete (per 50 ml) at 4 °C. The dialysate was sedimented for 5 min at 13,000 r.p.m., 14,926g, and the supernatant was quick-frozen in small aliquots in liquid N₂ and stored at –80 °C.

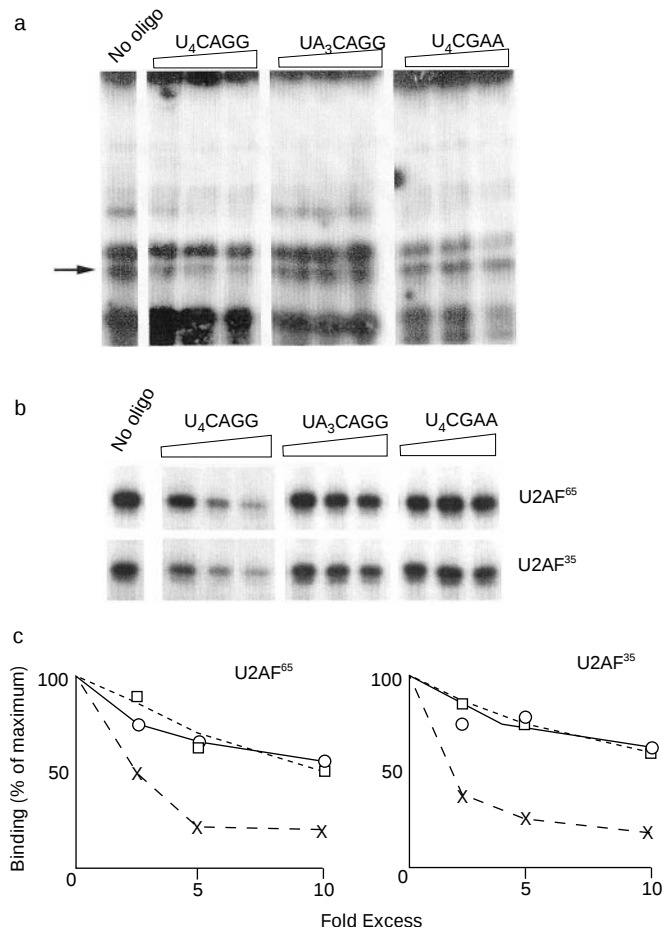


Figure 3 Crosslinking by ultraviolet, with competition. **a**, Transgenic extract was ultraviolet-crosslinked with ³²P-labelled U₄CAGG-containing RNA oligonucleotide in the presence of competitors. Increasing amounts of competitor oligonucleotides were in 2.5-fold, 5-fold and 10-fold molar excess over labelled RNA. **b**, Samples treated as in **a** were immunoprecipitated with U2AF⁶⁵ antibody. Both U2AF⁶⁵ and U2AF³⁵ were competed for specifically with U₄CAGG. **c**, Graphic representation of competition. Bands were quantified with the PhosphorImager system from Molecular Dynamics. Crosses, U₄CAGG; open squares, U₄CAGG; open circles, U₄CGAA.

Crosslinking with ultraviolet light

Embryonic extracts were incubated in the presence of buffer (20 mM HEPES-KOH, pH 7.6, 100 mM KCl, 0.2 mM EDTA, 10% glycerol, 0.5 mM DTT and 50 ng ml⁻¹ BSA) and 4 ng of 5' end ³²P-labelled RNA oligonucleotide for 20 min at 23 °C and transferred to a 96-well U-bottomed culture dish²⁰. Samples were irradiated with ultraviolet for 10 min on ice with a Mineralight lamp (Model UVG-54, UVP Inc.) with a 254-nm bulb 4 cm from the samples. The samples were subjected to electrophoresis on SDS-8% polyacrylamide gels, dried and exposed to X-ray film or phosphor screen. RNAs for ultraviolet crosslinking had the sequences shown in Fig. 2 preceded by ACUA and followed by CAC. Oligonucleotides were end-labelled with [α -³²P]ATP and purified on a Bio-Rad P-30 spin column.

Western blot analysis

Protein homogenates obtained by boiling worms, embryonic extracts or immunoprecipitates were subjected to electrophoresis and Western blotting as previously described¹².

Immunodepletion and immunoprecipitation

Extracts were diluted to 5 mg ml⁻¹, and 50- μ l aliquots of Protein-A-Sepharose beads coupled with rabbit anti-U2AF⁶⁵ antibody¹² or non-immune sera were added to 200 μ l of extract. The mixtures were incubated for 16 h with rotation at 4 °C. The beads were sedimented and the supernatant was quick-frozen in liquid N₂ and stored at -80 °C. For immunoprecipitation, samples crosslinked with ultraviolet were incubated with beads coupled with anti-U2AF⁶⁵ or non-immune sera, and with agarose beads coupled with monoclonal Flag peptide antibody (Anti-Flag M2 Affinity gel, Sigma), for 1 h, with shaking, at 23 °C. Samples were sedimented and beads were washed twice with PBS for U2AF⁶⁵ or non-immune sera and Tris-buffered saline, for Flag antibody. Beads were resuspended in SDS loading buffer and treated as above.

Received 20 July; accepted 21 October 1999.

- Moore, M. J., Query, C. C. & Sharp, P. A. in *The RNA world* (eds Gesteland, R. F. & Atkins, J. E.) 303–357 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993).
- Reed, R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113–2123 (1989).
- Ruskin, B., Zamore, P. D. & Green, M. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 201–219 (1988).
- Zamore, P. D. & Green, M. R. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.* **10**, 210–214 (1991).
- Zamore, P. D. & Green, M. Identification, purification and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl Acad. Sci. USA* **86**, 9243–9247 (1989).
- Zamore, P. D., Patton, J. G. & Green, M. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614 (1992).
- Valcarcel, J., Gaur, R. K., Singh, R. & Green, M. R. Interaction of U2AF⁶⁵ RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* **273**, 1706–1709 (1996).
- Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I. & Green, M. R. Cloning and intracellular localization of the U2 small ribonucleoprotein auxiliary factor small subunit. *Proc. Natl Acad. Sci. USA* **89**, 8769–8773 (1992).
- Rudner, D. Z., Kanaar, R., Breger, K. S. & Rio, D. C. Mutations in the small subunit of the *Drosophila* U2AF splicing factor cause lethality and developmental defects. *Proc. Natl Acad. Sci. USA* **93**, 10333–10337 (1996).
- Zorio, D. A. R. & Blumenthal, T. U2AF³⁵ is encoded by an essential gene clustered in an operon with RRM/cyclophilin in *C. elegans*. *RNA* **5**, 487–494 (1999).
- Blumenthal, T. & Steward, K. in *C. elegans II* (eds Riddle, D., Blumenthal, T., Meyer, B. & Priess, J.) 117–145 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997).
- Zorio, D. A. R., Lea, K. & Blumenthal, T. Cloning of *Caenorhabditis* U2AF⁶⁵: an alternative spliced RNA containing a novel exon. *Mol. Cell. Biol.* **17**, 946–953 (1997).
- Rudner, D. Z., Breger, K. S., Kanaar, R., Adams, M. D. & Rio, D. C. RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high affinity binding. *Mol. Cell. Biol.* **18**, 4004–4011 (1998).
- Conrad, R., Liou, R. F. & Blumenthal, T. Functional analysis of a *C. elegans* trans-splice acceptor. *Nucleic Acids Res.* **21**, 913–919 (1993).
- Zhang, H.-B. & Blumenthal, T. Functional analysis of an intron 3' splice site in *Caenorhabditis elegans*. *RNA* **2**, 380–388 (1996).
- Jones, D., Russnak, R. H., Kay, R. J. & Candido, E. P. Structure, expression, and evolution of a heat shock gene locus in *Caenorhabditis elegans* that is flanked by repetitive elements. *J. Biol. Chem.* **261**, 2006–2015 (1986).
- Liou, R.-F. & Blumenthal, T. Trans-spliced *Caenorhabditis elegans* mRNAs retain trimethylguanosine caps. *Mol. Cell. Biol.* **10**, 1764–1768 (1990).
- MacMorris, M. A., Zorio, D. A. R. & Blumenthal, T. An exon that prevents transport of a mature mRNA. *Proc. Natl Acad. Sci. USA* **96**, 3813–3818 (1999).
- Strocher, V. L. et al. DNA-protein interactions in the *Caenorhabditis elegans* embryo: oocyte and embryonic factors that bind to the promoter of the gut-specific *ges-1* gene. *Dev. Biol.* **163**, 367–380 (1994).
- Wilusz, J. & Shenk, T. A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. *Cell* **52**, 221–228 (1988).

Acknowledgements

We thank M. Green and T. Nilsen for sharing results before publication; I. Perez for helpful discussions; and M. MacMorris, A. Deshpande, D. Evans and D. Bentley for critical reading of the manuscript. This research was supported by the National Institute of General Medical Sciences.

Correspondence and requests for materials should be addressed to T.B. (e-mail: tom.blumenthal@uchsc.edu).

Inhibition of *msl-2* splicing by Sex-lethal reveals interaction between U2AF³⁵ and the 3' splice site AG

Livia Merendino, Sabine Guth, Daniel Bilbao, Concepción Martínez & Juan Valcárcel

Gene Expression Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

The protein Sex-lethal (SXL) controls dosage compensation in *Drosophila* by inhibiting the splicing and translation of *male-specific-lethal-2* (*msl-2*) transcripts^{1–6}. Here we report that splicing inhibition of *msl-2* requires a binding site for SXL at the polypyrimidine (poly(Y)) tract associated with the 3' splice site, and an unusually long distance between the poly(Y) tract and the conserved AG dinucleotide at the 3' end of the intron. Only this combination allows efficient blockage of U2 small nuclear ribonucleoprotein particle binding and displacement of the large subunit of the U2 auxiliary factor (U2AF⁶⁵) from the poly(Y) tract by SXL. Crosslinking experiments with ultraviolet light indicate that the small subunit of U2AF (U2AF³⁵) contacts the AG dinucleotide only when located in proximity to the poly(Y) tract. This interaction stabilizes U2AF⁶⁵ binding such that SXL can no longer displace it from the poly(Y) tract. Our results reveal a novel function for U2AF³⁵, a critical role for the 3' splice site AG at the earliest steps of spliceosome assembly and the need for a weakened U2AF³⁵–AG interaction to regulate intron removal.

Previous results *in vitro* indicated that interaction between SXL and intronic U-rich sequences was required for the inhibition of *msl-2* splicing⁶. But the introduction of these sequences at equivalent positions (immediately downstream of the 5' splice site and at the poly(Y) tract) within an adenovirus pre-mRNA (AdML) was not sufficient to make splicing of this RNA sensitive to SXL (Fig. 1a, compare lanes 1–3 and 4–6). The minimal set of *msl-2* sequences sufficient to confer regulation was the combination of the U₁₆ poly(Y) tract (which constitutes a high-affinity binding site for SXL) and the downstream sequence between the poly(Y) tract and the conserved AG dinucleotide at the 3' end of the intron (Fig. 1a, lanes 7–9).

Surprisingly, the primary sequence of the segment between the poly(Y) tract and AG was not important, because mutation of each nucleotide, or replacement by an heterologous sequence of the same length (Fig. 1b, β (3)m), did not affect *msl-2* regulation by SXL (Fig. 1a, lanes 10 and 11, and data not shown). Replacement by a shorter sequence present between the poly(Y) tract and AG in AdML (Fig. 1b, A(3)m) resulted in a substantial loss of regulation (Fig. 1a, lanes 12–14). These results indicate that the length of the segment between the poly(Y) tract and AG is critical for SXL-mediated intron retention. Consistent with this hypothesis is the observation that mutation of the AG dinucleotide in A(3)m RNA (Fig. 1b, long A(3)m), such that a further downstream AG was used, restored regulation (Fig. 1a, compare lanes 12–14 and 15–17). Conversely, the introduction of an AG dinucleotide four nucleotides downstream of the *msl-2* poly(Y) tract (Fig. 1b, short (3)m) substantially decreased regulation (Fig. 1a, compare lanes 1–3 and 18–20). The effect of reducing the poly(Y)-tract–AG distance on SXL regulation was also observed *in vivo* in *Drosophila* Schneider cells transiently transfected with *msl-2* or A(3)m reporter plasmids (Fig. 1c).

Two models could explain these observations. First, the poly(Y)-tract–AG distance could influence SXL binding; but our experiments failed to support this possibility (data not shown). Alternatively, an unusually long poly(Y)-tract–AG distance could weaken

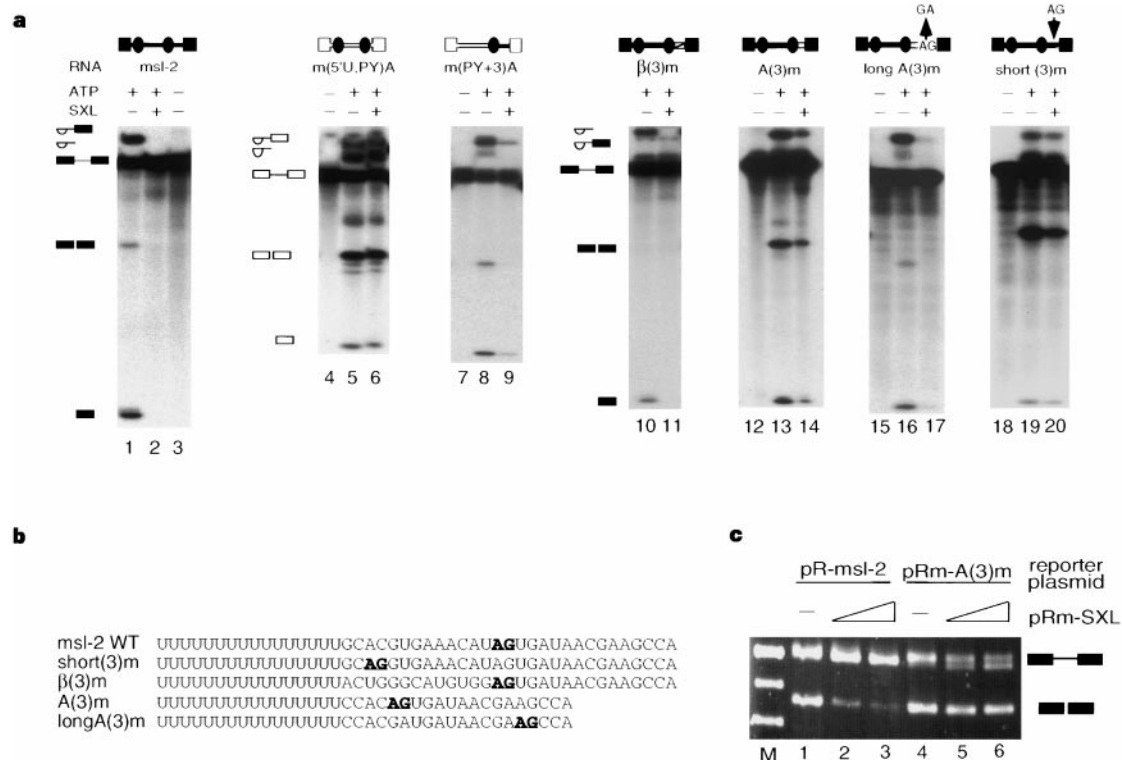


Figure 1 An unusually long poly(Y)-tract-AG distance is required for *msl-2* splicing regulation by SXL. **a**, Mapping *msl-2* sequences conferring SXL sensitivity to adenovirus (AdML) pre-mRNA splicing. Splicing mixtures containing the indicated radioactively labelled RNAs were incubated in the absence or presence of ATP or SXL (200 ng μL^{-1}) and the RNA was fractionated on denaturing polyacrylamide gels. *msl-2* sequences are shown in black and AdML sequences in white; the ovals represent SXL-binding sites. m(X)A indicates that sequence X from *msl-2* replaces the sequence at the corresponding position in AdML pre-mRNA. 5'U indicates a stretch of 11 U bases adjacent to the 5' splice site and PY represents a stretch of 16 U bases present at the poly(Y) tract of *msl-2*.

'3' indicates the 13-nucleotide sequence between the poly(Y) tract and AG at the 3' end of the *msl-2* intron. **b**, Primary sequence of the 3'-splice-site region of *msl-2* and mutant derivatives. **c**, The poly(Y)-tract-AG distance influences SXL regulation *in vivo* in *Drosophila* cells. Cytoplasmic RNA was isolated from Schneider cells transfected with reporter plasmids (20 ng cm^{-2}) and increasing concentrations (20 and 53 ng cm^{-2}) of an SXL expression plasmid, and analysed as described previously⁶. The amplification products corresponding to unspliced and spliced RNAs are indicated. M, molecular weight markers (400, 500 and 600 base pairs).

the 3'-splice-site recognition such that only in this situation could SXL interfere with the assembly or the activation of the splicing apparatus. To test this possibility, we analysed 3' splice-site-recognition by the U2 small nuclear ribonucleoprotein particle (snRNP) by using nuclear extracts in which other steps in spliceosome assembly were inhibited by ribonuclease (RNase)-H-mediated digestion of the 5' end of U1 small nuclear RNA (Fig. 2a). U2 snRNP binding was less efficient in wild-type *msl-2* than in the mutant containing a shorter poly(Y)-tract-AG distance (Fig. 2a, compare lanes 1 and 3). In addition, U2 snRNP binding was more efficiently blocked by SXL when the poly(Y)-tract-AG distance was longer (Fig. 2a, compare lanes 2 and 4). These results are consistent with a model in which a slower or less stable assembly of U2 snRNP on substrates containing a long poly(Y)-tract-AG distance allows regulation by SXL.

To analyse the molecular mechanism behind these observations, we examined whether SXL could differentially affect the binding of U2AF in a poly(Y)-tract-AG-distance-dependent manner. U2AF binding was assessed by crosslinking with ultraviolet light, followed by immunoprecipitation. U2AF is composed of two subunits of relative molecular mass 65,000 and 35,000 (M_r 65K and 35K)⁷. Whereas U2AF⁶⁵ recognizes the poly(Y) tract⁸, the function of U2AF³⁵ is not well understood^{9,10}. The results shown in Fig. 2b indicate that, whereas the presence of SXL decreased U2AF⁶⁵ crosslinking to *msl-2* RNA (lanes 1–3), it did not significantly decrease the crosslinking to a mutant substrate with a short poly(Y)-tract-AG distance (short (3)m, lanes 4–6). These observations suggest that a short poly(Y)-tract-AG distance makes U2AF⁶⁵ binding resistant to SXL challenge. Previous analyses have shown that the presence of an AG dinucleotide does not affect RNA binding by

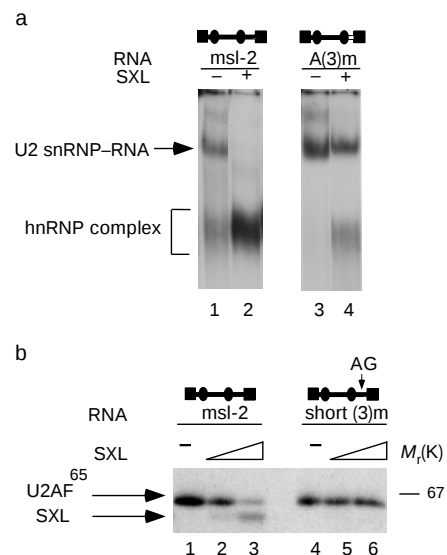


Figure 2 The poly(Y)-tract-AG distance affects early events in 3'-splice-site recognition. **a**, Assembly of U2 snRNP on the indicated RNAs was analysed by electrophoresis on non-denaturing gels of reaction mixtures set up as for Fig. 1 by using U1 inactivated nuclear extracts, in the presence or absence of SXL (200 ng μL^{-1}). **b**, Displacement of U2AF⁶⁵ by SXL was analysed by ultraviolet light irradiation and immunoprecipitation with anti-U2AF⁶⁵ antibodies of splicing mixtures containing RNAs uniformly labelled with [α -³²P]UTP, in the absence or presence of SXL (50 and 200 ng μL^{-1}). The crosslinked proteins were fractionated in SDS-polyacrylamide gels. The positions of U2AF⁶⁵ and recombinant SXL, which is also partly precipitated by the antibodies, are indicated.

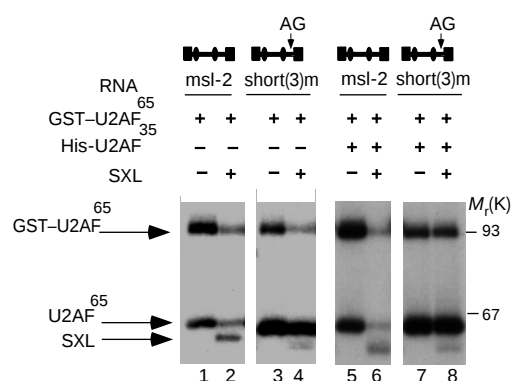


Figure 3 U2AF³⁵ makes U2AF⁶⁵ binding refractory to inhibition by SXL. Ultraviolet crosslinking and immunoprecipitation was performed as in Fig. 2b on splicing mixtures containing 20 ng μl^{-1} recombinant glutathione *S*-transferase (GST)–U2AF⁶⁵ (93K), without or with 7 ng μl^{-1} recombinant His–U2AF³⁵, using either *msl-2* or short (3)m RNAs uniformly labelled with [α -³²P]UTP in the absence or presence of 50 ng μl^{-1} SXL.

purified U2AF⁶⁵ (refs 8 and 11). Consistent with this is the observation that the displacement by SXL of recombinant glutathione *S*-transferase–U2AF⁶⁵ added to nuclear extracts was similar in both substrates (Fig. 3, lanes 1–4). These results could be explained if additional factors, probably associated with endogenous U2AF⁶⁵ and not available to recombinant protein added to extracts, were to stabilize U2AF⁶⁵ binding when the poly(Y)-tract–AG distance was short.

To identify factors that can differentially recognize the AG dinucleotide depending on its distance to the poly(Y) tract, three RNAs were synthesized (Fig. 4a): *msl-2*, containing a single radioactive phosphate between the A and the G at the 3' splice site; a mutant *msl-2* RNA containing a single radioactive phosphate between the A and the G of an AG dinucleotide located four nucleotides downstream of the poly(Y) tract ($\Delta(3)\text{m}$); and a derivative of the previous substrate in which the AG was mutated to CA and contained a single radioactive phosphate between these two nucleotides ($\Delta(3)\text{m-CA}$). Incubation of these RNAs with nuclear extracts under splicing conditions and ultraviolet irradiation

identified a polypeptide of 35K that was crosslinked to the $\Delta(3)\text{m}$ substrate but not to *msl-2* (Fig. 4b, lanes 1 and 2). Significantly, this crosslinked species was immunoprecipitated by antibodies against U2AF³⁵ (Fig. 4b, lane 4). To confirm the identity of the 35K crosslink and also to test whether the recognition of AG was specific, experiments were performed with $\Delta(3)\text{m}$ and $\Delta(3)\text{m-CA}$ RNAs. Figure 4c shows that the 35K protein was precipitated by both anti-U2AF³⁵ antibodies (lane 3) and, as expected from the extremely tight association between the two subunits^{8,12}, by anti-U2AF⁶⁵ antibodies (lane 5), but not by a control antibody (lane 7). Crosslinking of this polypeptide to the $\Delta(3)\text{m-CA}$ mutant was significantly lower (lanes 2, 4 and 6). A similar crosslinking pattern was observed with purified recombinant U2AF⁶⁵–U2AF³⁵ (Fig. 4d). Taken together, these observations indicate that U2AF⁶⁵ targets U2AF³⁵ to interact specifically with the 3' splice site AG when located in close proximity to the poly(Y) tract.

The U2AF³⁵–AG interaction could stabilize the binding of U2AF⁶⁵ to the poly(Y) tract and make the displacement of U2AF⁶⁵ by SXL less efficient. A long poly(Y)-tract–AG distance, in contrast, could weaken the U2AF³⁵–AG interaction and allow regulation by SXL. The experiment shown in Fig. 3 tests this hypothesis by comparing the displacement by SXL of endogenous U2AF with the displacement of recombinant U2AF⁶⁵ in the absence or presence of recombinant U2AF³⁵. Whereas SXL displaced recombinant U2AF⁶⁵ similarly, regardless of the poly(Y)-tract–AG distance (Fig. 3, lanes 1–4), the presence of recombinant U2AF³⁵ resulted in the displacement of recombinant U2AF⁶⁵ only when the poly(Y)-tract–AG distance was longer (lanes 5–8), thus mimicking the behaviour of endogenous U2AF.

Taken together, the results of Figs 2–4 offer an explanation for the requirement of an unusually long poly(Y)-tract–AG distance for the SXL-mediated inhibition of *msl-2* splicing reported in Fig. 1: to prevent the additional stabilization of U2AF⁶⁵ binding conferred by the interaction between U2AF³⁵ and the AG dinucleotide. This feature is not required when regulation involves diverting splicing to an alternative 3' splice site, as in *transformer* pre-mRNA¹³ (L.M., S.G., D.B., C.M. and J.V., manuscript in preparation). The distance between splicing signals at the 3' splice site has been implicated in the efficiency of the second step of the splicing reaction in yeast^{14–16} and in mammals^{17–19}. In addition, introns with a weak poly(Y) tract have

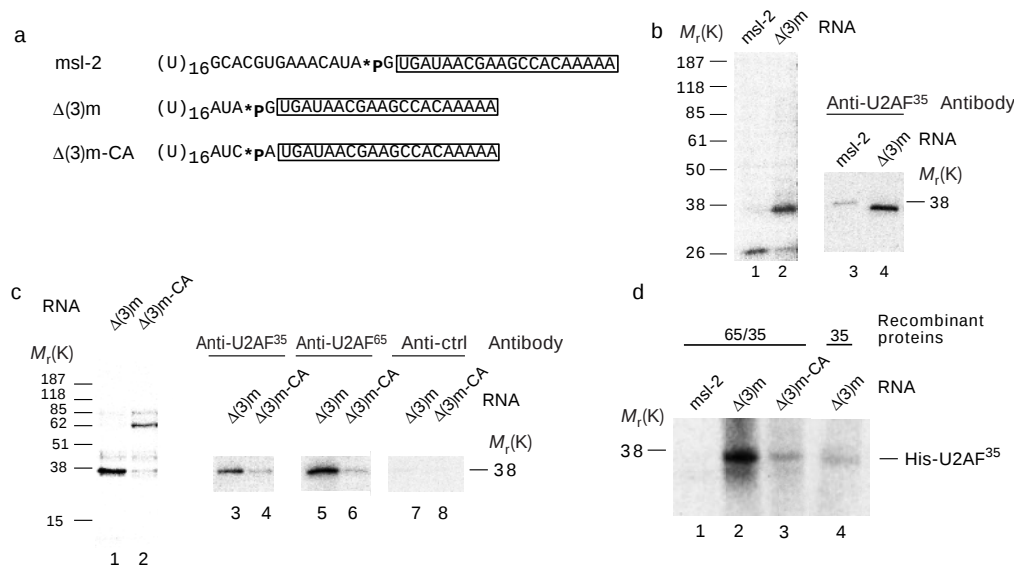


Figure 4 U2AF³⁵ contacts the 3' splice site AG in a distance-dependent manner. **a**, Site-specifically labelled RNA substrates. Radioactive phosphate is indicated by an asterisk. **b**, Detection of proteins crosslinked to site-specifically labelled *msl-2* and $\Delta(3)\text{m}$. After UV irradiation and digestion with RNase A and RNase T1, crosslinked polypeptides were either directly resolved on SDS–polyacrylamide gels or immunoprecipitated with anti-U2AF³⁵

antibodies. **c**, Products of UV crosslinking with $\Delta(3)\text{m}$ and $\Delta(3)\text{m-CA}$ RNAs (lanes 1 and 2) were immunoprecipitated with anti-U2AF³⁵, anti-U2AF⁶⁵ or control antibodies (anti-ctrl) before fractionation on SDS–polyacrylamide gels. **d**, Recombinant purified U2AF⁶⁵ (0.75 ng μl^{-1}) with or without U2AF³⁵ (1 ng μl^{-1}) were incubated with the indicated RNAs in the presence of yeast tRNA (0.3 ng μl^{-1}) and processed as in **b** (without immunoprecipitation).

been found to depend on the presence of a nearby AG dinucleotide to undergo the first step of splicing in mammals¹⁷. Our results show that even the earliest events in 3'-splice-site recognition are affected by the proximity between the poly(Y) tract and the AG dinucleotide. Furthermore, they show that the underlying molecular interactions also operate in RNAs with strong poly(Y) tracts highly enriched in U bases, and that these interactions can have a crucial role in splicing regulation, illustrating how the study of regulatory processes can illuminate basic mechanisms of gene expression. □

Methods

Constructs

pGEM-*msl-2*, pRm-*msl-2* and pRm-SXL were as described previously⁶, as was pAdML²⁰. All other clones were generated by PCR-based site-directed mutagenesis and confirmed by sequencing. In pm(5'U,PY)A, the stretches of 11 and 16 T bases corresponding to the SXL binding sites in *msl-2* RNA were introduced five nucleotides downstream of the 5' splice site and at the poly(Y) tract of the AdML intron, respectively. pm(PY + 3)A was prepared by replacing positions 107–119 in the AdML intron²⁰ by positions 103–131 in the *msl-2* intron¹. pβ(3)m was generated by replacing positions 119–131 in the *msl-2* intron by the human β-globin sequence 5'-ACTGGGCATGTGG-3'. pA(3)m was obtained by replacing positions 119–131 in the *msl-2* intron by positions 116–119 in the AdML intron. pshort(3)m was obtained by mutating position 122 of *msl-2* intron to G. pRm-A(3)m was generated by recloning the insert of pA(3)m into pRmHa-3 (ref. 6).

Splicing assays and RNA analysis

Transcription *in vitro*, splicing, spliceosome assembly and RNA analysis were as described^{6,7}.

Extracts and purified proteins

Nuclear extracts were prepared from HeLa cells as described²¹. RNase-H-mediated inactivation of U1 snRNP was performed by incubating 50 μl of extract with 500 pmol of an oligodeoxynucleotide complementary to the first 15 nucleotides of U1 snRNA in the presence of 1.1 mM ATP, 5.6 mM creatine phosphate, 2.2 mM MgCl₂, 40 U RNasin and 3 U RNase H at 37 °C for 15 min and at 30 °C for 45 min. Recombinant purified SXL and U2AF⁶⁵ were prepared as described^{8,13}. Recombinant purified U2AF³⁵ was purified from baculovirus-infected cells as described⁹.

Site-specifically labelled pre-mRNAs

These were obtained as described²² by using chemically synthesized 3' RNAs 5' end-labelled with [γ-³²P]ATP and T₄ polynucleotide kinase.

Crosslinking by ultraviolet light and immunoprecipitation

These were performed as described¹⁰ on 27-μl splicing mixtures by using 35 μl of anti-U2AF⁶⁵ hybridoma supernatant, 20 μl of anti-U2AF³⁵ (ref. 9) or control anti-β-importin antisera. The precipitates were washed with 500 mM NaCl (for U2AF⁶⁵) or 100 mM NaCl (for U2AF³⁵) buffers, and the bound proteins were fractionated on SDS–10% polyacrylamide gels.

Received 27 July; accepted 5 November 1999.

1. Bashaw, G. & Baker, B. The *msl-2* dosage compensation gene of *Drosophila* encodes a putative DNA-binding protein whose expression is sex specifically regulated by *Sex-lethal*. *Development* **121**, 3245–3258 (1995).
2. Kelley, R. *et al.* Expression of *Msl-2* causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell* **81**, 867–877 (1995).
3. Zhou, S. *et al.* Male-specific lethal 2, a dosage compensation gene of *Drosophila*, undergoes sex-specific regulation and encodes a protein with RING finger and a metallothionein-like cysteine cluster. *EMBO J.* **14**, 2884–2895 (1995).
4. Bashaw, G. & Baker, B. The regulation of the *Drosophila msl-2* gene reveals a function for *Sex-lethal* in translational control. *Cell* **89**, 789–798 (1997).
5. Kelley, R., Wang, J., Bell, L. & Kuroda, M. *Sex-lethal* controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* **387**, 195–199 (1997).
6. Gebauer, F., Merendino, L., Hentze, M. & Valcárcel, J. The *Drosophila* splicing regulator *Sex-lethal* directly inhibits translation of male-specific-lethal 2 mRNA. *RNA* **4**, 142–150 (1998).
7. Zamore, P. & Green, M. Identification, purification and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl Acad. Sci. USA* **86**, 9243–9247 (1989).
8. Zamore, P., Patton, J. & Green, M. R. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614 (1992).
9. Zuo, P. & Maniatis, T. The splicing factor U2AF³⁵ mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* **10**, 1356–1368 (1996).
10. Kan, J. & Green, M. R. Pre-mRNA splicing of IgM exons M1 and M2 is directed by juxtaposed splicing enhancer and inhibitor. *Genes Dev.* **13**, 462–471 (1999).
11. Singh, R., Valcárcel, J. & Green, M. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173–1176 (1995).
12. Zamore, P. D. & Green, M. R. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.* **10**, 207–214 (1991).
13. Valcárcel, J., Singh, R., Zamore, P. & Green, M. R. The protein *Sex-lethal* antagonizes the splicing factor U2AF to regulate alternative splicing of *transformer* pre-mRNA. *Nature* **362**, 171–175 (1993).
14. Frank, D. & Guthrie, C. An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev.* **6**, 2112–2124 (1992).

15. Umen, J. G. & Guthrie, C. The second catalytic step of pre-mRNA splicing. *RNA* **1**, 869–885 (1995).
16. Luukkainen, B. G. M. & Séraphin, B. The roles of branchpoint-3' splice site spacing and interaction between intron-terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. *EMBO J.* **16**, 779–792 (1997).
17. Reed, R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113–2123 (1989).
18. Smith, C. W. J., Chu, T. T. & Nadal-Ginard, B. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell. Biol.* **13**, 4939–4952 (1993).
19. Chiara, M. D., Palandjian, L., Feld Kramer, R. & Reed, R. Evidence that U5 snRNP recognizes the 5' splice site for catalytic step II in mammals. *EMBO J.* **16**, 4746–4759 (1997).
20. Zapp, M. L. & Berget, S. M. Evidence for nuclear factors involved in recognition of 5' splice sites. *Nucleic Acids Res.* **17**, 2655–2674 (1989).
21. Dignam, J., Lebovitz, R. & Roeder, R. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
22. Moore, M. & Sharp, P. Site-specific modification of pre-mRNA: the 2' hydroxyl groups at the splice sites. *Science* **256**, 992–997 (1992).

Acknowledgements

We thank S. Wu and M. R. Green for communicating results before publication, and for experimental advice and discussions; T. Maniatis for anti-U2AF³⁵ antibodies and the U2AF³⁵ recombinant baculovirus construct; and B. Séraphin and I. Mattaj for advice.

Correspondence and requests for materials should be addressed to J.V. (e-mail: juan.valcarcel@embl-heidelberg.de).

correction

Long-range electrostatic attraction between like-charge spheres in a charged pore

W. Richard Bowen & Adel O. Sharif

Nature **393**, 663–665 (1998)

In Fig. 3, an incorrect file was plotted for the data in the inset for the isolated spheres (square symbols). This is now plotted correctly, together with a more informative caption. This correction does not affect the conclusions of our paper. □

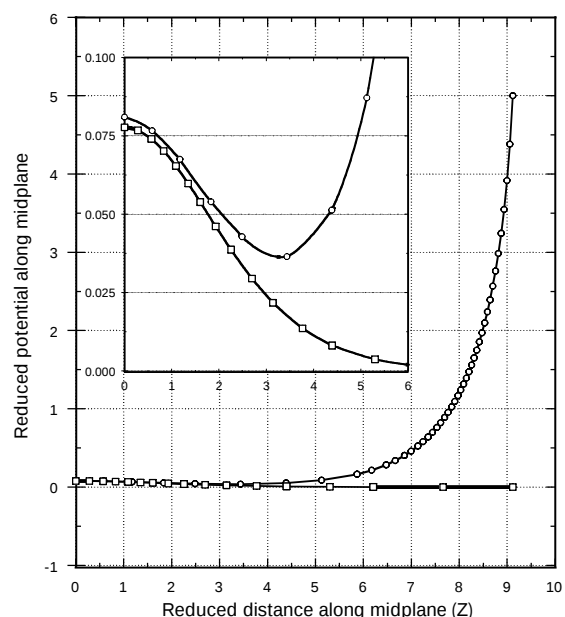


Figure 3 Potential along the midplane between two spheres at reduced surface–surface separation distance ($L = 6.0$): isolated spheres (square symbols) and spheres confined in a pore (circles). Zero distance corresponds to the line of centres of the spheres. On the right-hand side of the main graph, the pore wall is reached at $\Psi = 5.0$.